

Resumo: Neste artigo analisam-se dois casos práticos de produção informacional considerados prioritários no contexto da estruturação do repositório digital confiável de uma instituição da administração pública portuguesa. Um estudo que se insere na área científica da Ciência da Informação, mais concretamente no campo da Gestão da Informação (GI), assumindo-se a função de preservação da informação como sua variável.

Aborda-se, por um lado, a adoção da estratégia de preservação de normalização de formatos a ingerir no repositório – através do formato PDF/A –, um trabalho que tem vindo a ser realizado na instituição nos últimos anos, e por outro lado, as tarefas e especificações a desenvolver no âmbito da gestão e preservação da produção informacional comunicada via *email*. Caracteriza-se o formato PDF, as suas variantes e contextos de aplicação, assim como se define e descreve o funcionamento do *email* e o seu ciclo de vida, salientando a importância que lhe deverá ser atribuída na GI. Uma abordagem que envolve o desenvolvimento de políticas de gestão de informação e o foco na gestão integrada da informação em todo o seu ciclo de vida, assumindo a sua preservação numa perspetiva de acesso e uso continuado, só possível se for efetivamente integrada nos processos de gestão do sistema de informação organizacional, desde a fase de produção, receção ou captura.

Apresentam-se diferentes perspetivas, casos de boas práticas, bem como exemplos de orientações e ferramentas disponíveis.

Palavras-chave: Ciência da Informação; Gestão da Informação; Preservação da Informação; Gestão de correio eletrónico

Abstract: This paper presents some results of the analysis of two practical cases of informational production considered as a priority in the context of trustworthy digital repository structuration in a Portuguese public administration institution. This is an Information Science (IS) study in the field of Information Management (IM), considering the function of information preservation as an IM variable under a systemic perspective.

It addresses, on the one hand, the adoption of formats' standardization as a preservation strategy based on the adoption of PDF/A, a work that has been carried out in recent years in the institution, and on the other hand, tasks and specifications that should be developed regarding the management and preservation of informational production communicated by email. It is characterized PDF format, variants and application contexts, and it is defined and described how works the technology of email, the need and importance to be considered in IM processes. An approach that involves the development of an email management policy, the email management process throughout its life cycle, assuming its preservation in a perspective of continuous access and use, only possible if it's effectively integrated into the management processes of organizational information system, from the stage of production, reception or capture.

It presents different perspectives, good practices and reference guidelines and tools.

Keywords: Information Science; Information Management; Information Preservation; Email management

Introdução¹

A informação é, hoje, crescentemente produzida em meio digital e, quando tal não se verifica, procede-se à chamada “desmaterialização”, isto é, à substituição da tramitação física pela tramitação digital, promovendo a sua digitalização.

Desta forma, o sujeito humano deixa de poder aceder diretamente a informação, envolvendo, agora, novas formas de estruturação e codificação das representações mentais - ou mentefactos - e da sua fixação para acesso e utilização futura (Pinto, 2013).

Com o meio digital, a preservação da informação começa, desde logo, na conceção e implementação da plataforma tecnológica na qual será produzida, recebida, capturada, organizada, armazenada e difundida a informação. Serão, assim, considerados o *hardware* e *software* utilizados, os formatos adotados, a recolha atempada da metainformação administrativa, técnica, estrutural, descritiva ou de preservação que permita a sua futura referência e o desencadear das várias estratégias de preservação que agirão sobre as diferentes dimensões da unidade informacional. Inserem-se, aqui, quer a bidimensionalidade do documento analógico, quer a pluridimensionalidade do “objeto” digital e a assunção de um papel que não poderá ser imputado a “informáticos”, mas que convoca permanentemente parcerias entre o gestor da informação e os diferentes atores, informáticos incluídos (PINTO, 2014).

O desafio que se nos coloca é, pois, o de fazer face à *Gestão da Informação* (GI) em meio digital (PINTO, 2005, 2009), parte integrante de *Sistemas de Informação* tendencialmente híbridos e nos quais a obsolescência tecnológica, visível a vários níveis (*hardware*, *software*, suportes de armazenamento, formatos, etc.), é um entre vários fatores a considerar. Configura-se um complexo processo integrado de gestão sustentado em conhecimento teórico, instrumentos, ferramentas e técnicas de operacionalização que estarão presentes ao longo de todo o ciclo de vida e de gestão da informação dando corpo ao SGIAP, isto é, ao *Sistema de Gestão de Informação Ativa e Permanente*².

Neste contexto, emerge a necessidade de estruturação de repositórios digitais confiáveis que tem despertado cada vez mais atenções nos últimos anos, sendo a certificação do repositório um dos objetivos a atingir no âmbito da Gestão da Informação em contexto organizacional. É urgente garantir o armazenamento seguro, o acesso controlado e a preservação da informação (documentos digitalizados e nado-digitais) no longo termo e,

¹ Este artigo aborda duas das vertentes estudadas no âmbito da dissertação defendida publicamente na FEUP em julho de 2014: OLIVEIRA, Hugo Azevedo de - *A Preservação da Informação: um contributo para a implementação de um arquivo digital certificável no Município do Porto*. Porto: Faculdade de Engenharia da Universidade do Porto, 2014. Dissertação de Mestrado em Ciência da Informação. Orientadora: Maria Fernanda Martins; coorientadora: Maria Manuela Pinto.

² Ver: Gavina, 2014 e Oliveira, 2014.

assim, corresponder à procura interna por parte dos prosumidores que constituem o universo de colaboradores da organização/instituição e respetivos serviços, bem como à procura externa de informação, cumprindo o requisito de disponibilização e partilha da informação, cada vez mais em foco nas políticas nacionais e internacionais desenvolvidas, nomeadamente, em torno dos *open data* e do *open government*.

No entanto, a concretização deste objetivo exige um minucioso e complexo trabalho a montante, sobretudo no que respeita à produção informacional, que se confronta sucessivamente com inovações tecnológicas que, se por um lado oferecem novas soluções e por outro suscitam novos problemas.

No que respeita ao caso estudado – a Câmara Municipal do Porto (CMP) –, identificaram-se ao nível da gestão da fase de produção informacional dois tópicos que, pelo seu impacto, foram considerados prioritários: a adoção do formato PDF e o controlo dos *emails*³.

A adoção do formato PDF impôs-se nos últimos anos, dadas as características de interoperabilidade e de preservação que se lhe reconhecem, tendo-se apresentado para a instituição uma clara opção a dois níveis: 1) no que respeita à adoção da estratégia de normalização de formatos durante ou no final da tramitação administrativa, que permite que hoje estejam já assegurados os requisitos para a preservação e acesso a longo prazo de grande parte da produção informacional atual da CMP; 2) e ao nível da digitalização retroativa da informação acumulada, realizada desde o início da entrada em funcionamento da Unidade Central de Digitalização (UCD), em interação com o Gabinete do Município e serviços autárquicos (FERNANDES; BRANDÃO; CoOSTA, 2010; BRANDÃO, 2010; CMP, 2012). Não obstante, a evolução tecnológica tornou necessário analisar o estado da arte e apontar novas perspetivas, face aos últimos desenvolvimentos neste domínio (SOUSA, 2013; CMP, 2012; OLIVEIRA, 2014).

Por sua vez, o domínio e volume alcançado pela informação comunicada via *email* requer um controlo que passa pela compreensão do funcionamento deste meio de comunicação e pela identificação dos problemas que coloca, por forma a evidenciar a importância que lhe deverá ser atribuída no âmbito da GI e os requisitos requeridos para a sua efetivação. Uma perspetivação que envolve o desenvolvimento de uma *Política de Gestão de Email*, o *processo de gestão do email* ao longo de todo o seu ciclo de vida, assumindo a sua preservação numa perspetiva de acesso e uso continuado no longo prazo, só possível se for efetivamente integrada nos processos de gestão do *Sistema de Informação*

³ Da bibliografia produzida sobre os projetos mais recentes da CMP destacam-se: FERNANDES; BRANDÃO; COSTA, 2010; BRANDÃO, 2010; CMP, 2012; SOUSA, 2013; OLIVEIRA, 2014).

Organizacional (SIO), da produção/captura à preservação no longo prazo (BRANDÃO, 2010; SOUSA, 2013; CMP, 2012; OLIVEIRA, 2014).

Esta é, pois, uma reflexão que aborda apenas dois tópicos, dos muitos que o desafio da preservação da informação em meio digital coloca no topo das prioridades de GI em instituições e organizações, contribuindo para a concretização de dois objetivos indissociáveis: a elaboração do *Plano de Preservação da Informação* e a estruturação de um *Repositório Digital Confiável*, que deverão existir qualquer que seja a instituição, a organização ou setor de atividade.

1. A GI e a Preservação

A abordagem a desenvolver foca a gestão do “produto informacional”, conscientes que, “numa qualquer organização a preservação começa, desde logo, com a análise dos contextos de produção da informação, dos seus fluxos, do seu uso, dos meios de armazenamento e acesso, bem como dos instrumentos de controlo existentes (planos de classificação, linguagens de indexação, controle de autoridade nos pontos de acesso, etc.), do processo de automação desenvolvido e a desenvolver, apontando para a gestão integrada do ciclo de vida da informação (integrando, por exemplo, o *software* aplicacional e de suporte com o *software* de gestão do sistema de informação), o que terá repercussão na alteração dos processos de gestão da informação, atores e serviços responsáveis. Esta análise será essencial, por exemplo, para a especificação/validação dos requisitos de aquisição/desenvolvimento de aplicações informáticas ditas de Gestão Documental e de *Workflow*, bem como para a desmaterialização sustentada dos processos organizacionais” (PINTO, 2014).

O SIO afirma-se como conceito orientador sendo “constituído pelos diferentes tipos de informação registada ou não externamente ao sujeito [...], não importa qual o suporte (material e tecnológico), de acordo com uma estrutura (entidade produtora/recetora) prolongada pela ação na linha do tempo” (SILVA, 2006) e, numa perspetiva holística e sistémica, a GI procura assumir um único ciclo de gestão que acompanha todo o ciclo de vida da informação e, neste caso, todo o ciclo de vida do *email*.

A preservação da informação é perspectivada como uma variável da GI, estando, assim, presente em todo o ciclo de vida informacional, convocando as áreas da *Produção Informacional*, da *Organização e Representação da Informação* e do *Comportamento Informacional*, podendo ser considerada quer nos estudos científicos, quer na resolução de casos concretos e mantendo como objetivos garantir a autenticidade, a fiabilidade, a

integridade e a inteligibilidade da informação, bem como o acesso continuado no longo prazo.

Fig. 1 – Perspetiva informacional e científica (PINTO, 2014)



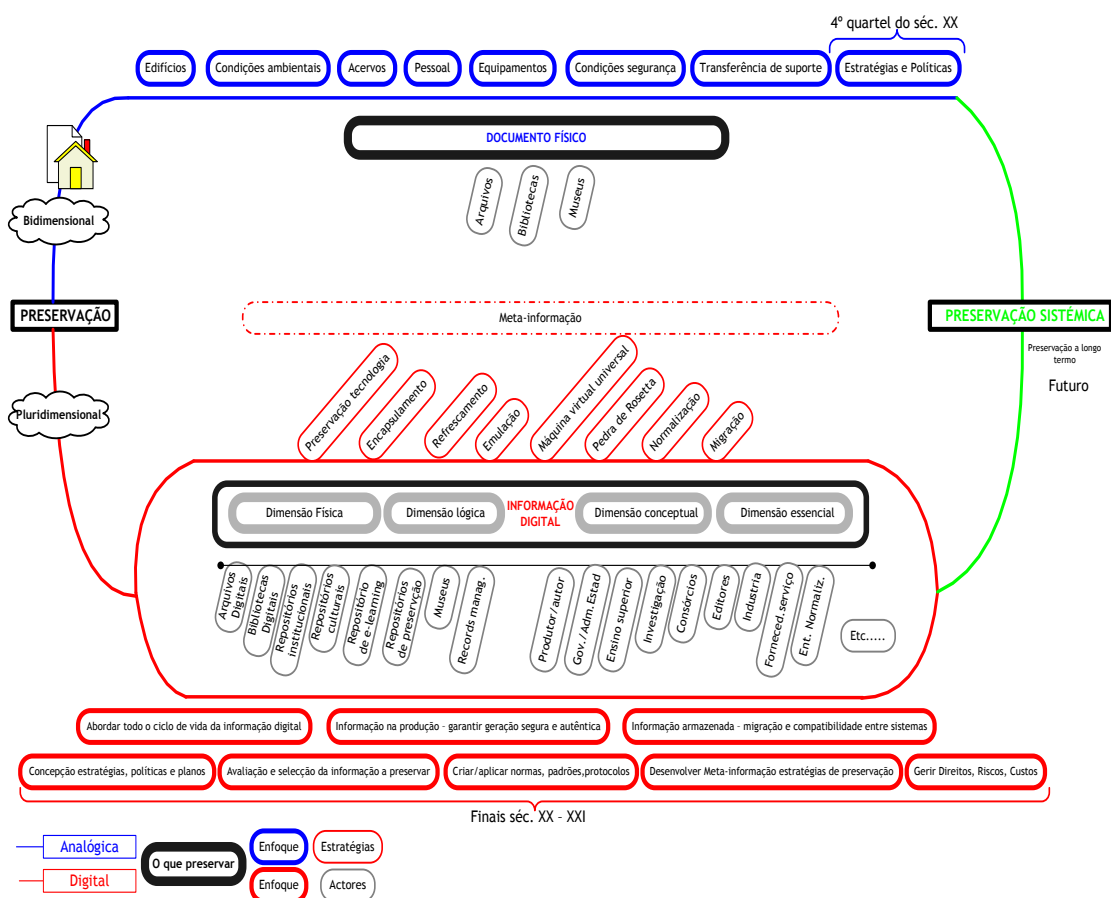
Como refere Pinto, ultrapassando posicionamentos que dicotomicamente enfatizavam ora a *dimensão informacional*, de difusão e de acesso, sobretudo ligados a bibliotecas, ora a *dimensão probatória e de armazenamento e preservação a longo prazo*, que colocava a questão da “autenticidade” no centro das preocupações da preservação dita digital, sobretudo em arquivos, ou que apontavam para o *pendor tecnológico*, temos decididamente que atender a todas as fases do ciclo de vida da informação e integrar o problema da preservação em meio digital no funcionamento corrente da organização, nas suas políticas e estratégias, bem como no quadro da implementação de modelos de processos e relações formais entre os diferentes atores no âmbito da economia/sociedade digital, apontando para a própria certificação dos repositórios, processos e serviços de GI (incluindo a *gestão da preservação*), solidamente alicerçados numa base teórico-metodológica que suportará a efetiva gestão do SIO (PINTO, 2009, 2014).

A informação a preservar deverá ser considerada na sua globalidade e complexidade, apelando a uma preservação sistémica aos vários níveis e entre os vários níveis, a saber: nível da *unidade informacional* (informação e metainformação); nível do *SIO* (seja digital, analógico ou híbrido); nível do *STI* (o Sistema Tecnológico de Informação que abarca toda a arquitetura computacional), nível da *Organização* (estrutura, processos, pessoas, SIO e STI; contexto interno e externo); nível *interorganizacional*; nível *nacional*; nível *global* (PINTO, 2009).

Também no caso da produção informacional via *email*, e independentemente desta integrar um SIO totalmente digital, híbrido ou mesmo analógico (via versão impressa), será o SIO e a missão/necessidades da Organização que o produz, acumula e usa, que norteia(m) a função preservação e os profissionais por ela responsáveis, não estando a sua

existência/configuração condicionada, por exemplo, por uma determinada tipologia de serviço de informação ou prévia classificação de carácter patrimonial, nem devendo ser consagrada qualquer distinção em termos de suporte como indiciado pela utilização do termo “Preservação Digital” (PINTO, 2014).

Fig. 2 - Um percurso que conduz à Preservação da Informação em sentido sistémico - PRESERVMAP (PINTO, 2009)

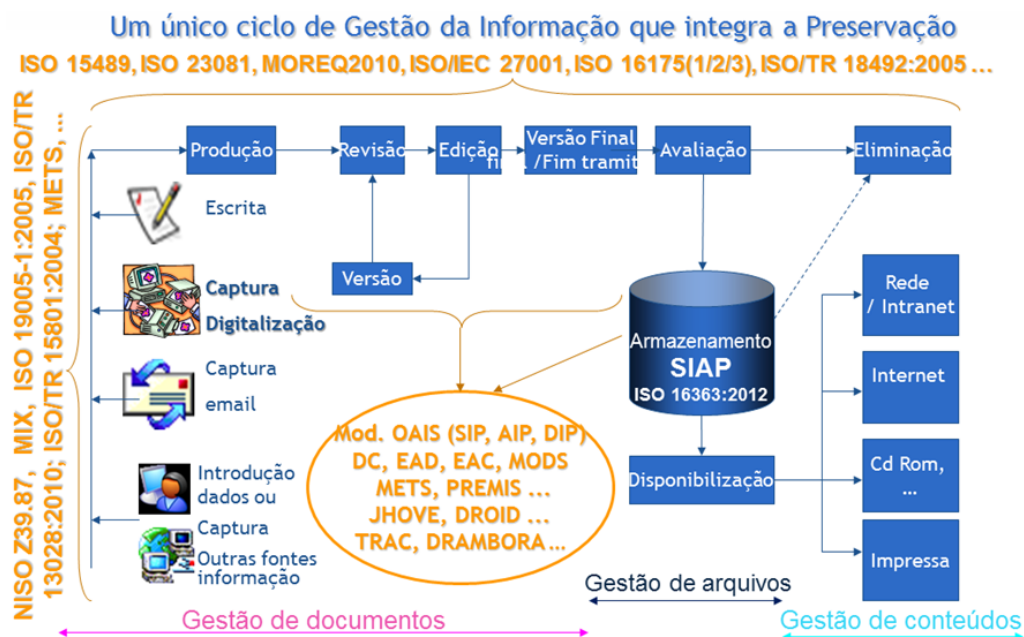


Significa, pois, que a preservação da informação assumida sistemicamente começa, desde logo, na conceção e implementação da plataforma tecnológica na qual será produzida, recebida, capturada, organizada, armazenada e difundida a informação, nomeadamente no *software* utilizado, nos formatos adotados, na recolha, na fase e no momento próprio, da metainformação administrativa, técnica, estrutural, descritiva ou de preservação que permita a sua futura referenciação e o desencadear das diferentes estratégias de preservação que agirão sobre as várias dimensões, seja a bidimensionalidade do documento analógico, seja a pluridimensionalidade do “objeto” digital. Um papel que não

poderá ser imputado aos “informáticos”, mas que convoca permanentemente parcerias entre o gestor da informação e os diferentes atores, informáticos incluídos (PINTO, 2014).

Como se constata na Fig. 3, desenvolve-se um complexo processo integrado de gestão que convoca conhecimento teórico, instrumentos, ferramentas e técnicas de operacionalização que estarão presentes ao longo de todo o ciclo de vida e gestão da informação, abrangendo momentos que, ainda hoje, são frequentemente assumidos de forma segmentada e sem articulação entre si e que deverão integrar o *Serviço de Gestão da Preservação da Informação*, tendo como referência a estrutura de serviços a desenvolver em torno da GI (OLIVEIRA, 2014; GAVINA, 2014).

Fig. 3 - Um único ciclo de gestão que integra a preservação (PINTO, 2014)



2. O formato PDF

A crescente necessidade de garantir a preservação e o acesso continuado à informação a longo prazo resulta, por um lado, da dificuldade de o ser humano deixar de poder aceder diretamente à informação, requerendo a mediação tecnológica e, por outro lado, porque o “objeto digital” é dinâmico, de acesso/comunicação assíncrona e multidirecionada, facilmente reutilizado/manipulado em grande escala, evidenciando a complexidade e a pluridimensionalidade inerente à produção informacional em meio digital.

A função de preservação da informação tem necessariamente que fazer frente aos perigos

que podem ameaçar qualquer uma das dimensões (Thibodeau, 2002) do processo de materialização da unidade informacional digital a saber:

- “a dimensão física (estamos perante registos, isto é, conjuntos de signos/símbolos gravados em suportes como discos, disquetes, CD’s ou DVD’s, muitas vezes residentes a milhares de quilômetros do ponto a partir do qual estão a ser acedidos, definindo o suporte a tipologia dos signos que têm que ser interpretáveis pelo hardware);
- a dimensão lógica (estamos perante um código compreensível para o computador - conjunto de informação organizada em estruturas de dados/formatos como .tiff, .pdf, .doc, etc.-, cuja existência depende num dado momento do registo físico, mas que não está vinculada a um suporte em particular);
- a dimensão conceptual (o código adquire um significado para o ser humano, o que não acontece na dimensão lógica ou física, isto é, os sinais digitais transformam-se em sinais analógicos e podem ser reconhecidos como o resultado apresentado – texto, imagem – e captado/(re)construído na mente do usuário/receptor);
- a dimensão essencial (o grupo de elementos essenciais que incorporam o propósito ou as características pelas quais se decidiu preservar a informação, incluindo, por ex. os aspectos relacionados com a autenticidade e o ciclo de vida, isto é, a metainformação administrativa, técnica, descritiva ou estrutural que garantirá a capacidade de a unidade informacional ser autodemonstrável, mantendo ligados a si os elementos contextualizadores da sua produção e ciclo de vida, sob o ponto de vista de contextual/ambiental – interno e externo –, informacional, orgânico, funcional e tecnológico)” (PINTO, 2014).

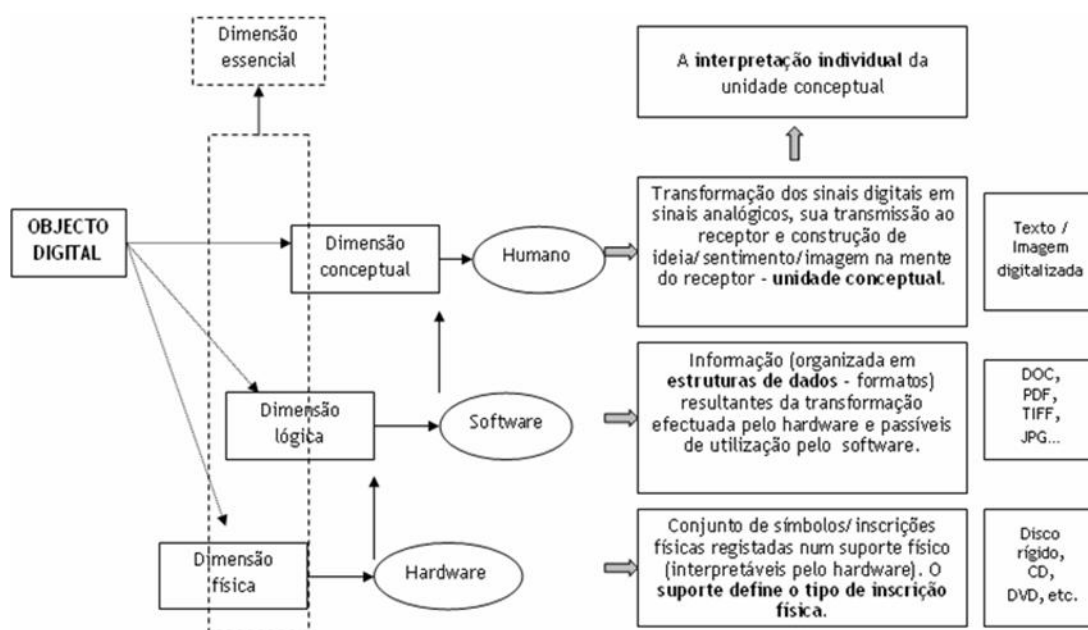
Neste contexto pluridimensional, a problemática dos *formatos* situa-se ao nível da preservação da dimensão lógica da unidade informacional digital obrigando a pensar a questão da obsolescência dos formatos, da necessidade de garantir a interoperabilidade, assim como o problema dos formatos proprietários, por forma a viabilizar o acionamento de estratégias como a da normalização de formatos, indicando os mais adequados em função das necessidades e propriedades da informação em causa.

Uma enunciação muito geral dos formatos permite-nos referenciar para além dos formatos nativos, o TIFF, o XML e o PDF sendo pertinente para a compreensão deste domínio a definição de *formato sustentável* que, de acordo com os *National Archives and*

Records Administration (NARA, USA) consiste na "[...] capacidade de aceder a um registo eletrónico em todo o seu ciclo de vida, independentemente da tecnologia utilizada quando foi originalmente criado" (FANNING, 2010). Entre as características de um formato sustentável encontram-se as seguintes:

- Documentação publicada e divulgação aberta;
- Adoção disseminada e uso;
- Formatos de autodescrição;
- Dependência externa;
- Impacto de Patentes;
- Mecanismo de proteção técnica.

Fig. 4 - Pluridimensionalidade a preservar (PINTO, 2009)⁴



a. O formato PDF (Portable Document Format)

O *Portable Document Format* (PDF) é, assim, um entre os diversos formatos mas diferencia-se pela garantia de interoperabilidade e de preservação no longo prazo,

⁴ Cf. Thibodeau, 2002.

exigindo que, para uma melhor eficiência e eficácia na sua adoção pelas organizações, se atenda ao contexto do seu aparecimento e especificidades desenvolvidas ao longo dos últimos anos.

O PDF é genericamente apresentado como um formato de ficheiro usado para apresentar os documentos de uma forma independente do *software* aplicacional, *hardware* e sistemas operativos. Cada ficheiro PDF encapsula uma descrição completa de um documento num *layout* fixo, incluindo o texto, fontes, gráficos e outras informações necessárias para a sua apresentação.

Surge como um formato proprietário, controlado pela Adobe, até que foi oficialmente lançado como um padrão aberto em julho de 2008 e publicado como ISO 32.000-1:2008, pela International Organization for Standardization⁵.

A sua rápida difusão resulta da necessidade de garantir que um documento nado-digital ou digitalizado seja não só visualizado de forma correta, mas também com uma versatilidade que ultrapassa as limitações de formatos como o TIFF, reduzindo o tamanho dos ficheiros e permitindo o armazenamento de objetos estruturados (texto, imagens vetoriais e imagens *bitmap*), a pesquisa no texto do documento e *Optical Character Recognition* (OCR) interno, assim como a inserção e classificação automática de metainformação (título, data de criação, autor, etc.).

Os níveis de conformidade e versões envolvem, entre outros, o PDF/A1, o PDF/A2, o PDF/A3, o PDF/X e o PDF/E, evidenciando uma evolução que decorre do facto de, apesar da generalizada aceitação do PDF, se verificar uma falta de uniformização resultante da diversidade das ferramentas usadas para a sua criação e deste formato não ser necessariamente *self-contained*, isto é, um formato em que as fontes necessárias para a visualização estivessem embebidas no próprio ficheiro. Estas limitações mantêm o PDF dependente de recursos externos diminuindo, conseqüentemente, a atratividade da sua utilização como formato que responda, nomeadamente, aos requisitos de preservação no longo prazo.

Para ultrapassar estas limitações são mobilizadas instituições como a AIIM (Association for Information and Image Management), a NEPS (National Printing Equipment Association), o Administrative Office of the U.S. Courts, e, posteriormente, a Library of Congress, os NARA, a Adobe e a Xerox, chegando-se a um formato para *arquivo a longo prazo*, o PDF/A-1, homologado como norma ISO 19.005-1:2005⁶.

Para Levenson (2008), o porquê de se escolher o formato PDF/A deve-se:

⁵ ISO 32000-1:2008 - *Document management : portable document format. Part 1: PDF 1.7.*

⁶ ISO 19005-1:2005 - *Document Management : electronic document file format for long term preservation. Part 1: use of PDF 1.4 (PDF/A-1).*

- à necessidade que resulta do negócio e o mercado consumidor estarem muitas vezes dessincronizados;
- a modelos com ciclos de atualização de três anos;
- à tentativa de emulação como “papel eletrônico”;
- à migração mínima;
- a uma apresentação precisa, consistente e previsível;
- à ausência de uma solução única para todos os problemas;
- ao leitor de livre utilização, multiplataforma e exigindo uma formação mínima.

O PDF/A apresenta-se, assim, como um formato de ficheiro normalizado e uma componente de uma estratégia de preservação abrangente, visando a resposta a três necessidades:

- Definir um formato de ficheiro que preserve a aparência visual estática de documentos eletrónicos ao longo do tempo;
- Fornecer uma estrutura para a gravação de metainformação sobre os documentos eletrónicos;
- Fornecer uma estrutura para definir a estrutura lógica e propriedades semânticas dos documentos eletrónicos (FANNING, 2010).

De acordo com a mesma autora, para existir uma boa apresentação/”renderização”, um documento PDF/A deve ter:

- Fontes e imagens incorporados;
- Elementos gráficos incorporados;
- Ausência de áudio ou de vídeo;
- Ausência de referências diretas ou indiretas a recursos externos;
- Metainformação;
- Ausência de elementos de segurança ou criptografia;
- Ausência de ficheiros incorporados (FANNING, 2010).

Na sua primeira versão – PDF/A-1 –, para além de ser um formato independente de qualquer plataforma (*software* ou *hardware*), é:

- autossuficiente, em termos de visualização e impressão (fontes embebidas, independência de cor relativamente à plataforma usada);
- autodescritivo (através da metainformação do PDF em formato XMP);
- e não está sob mecanismos de proteção e de restrição de acesso e disponibilização (livre uso da referência PDF e da especificação XMP) que

limitem o seu uso generalizado.

O PDF/A-1 suporta, ainda, a assinatura digital, desde que as fontes utilizadas estejam embebidas no formato.

Este formato não permite a encriptação, a compressão LZW (por motivos de direitos de propriedade), os ficheiros embebidos, as referências a conteúdos externos, transparências PDF, multimédia e JavaScript.

A norma especifica dois níveis de conformidade para os ficheiros PDF/A-1, aplicáveis aos casos de utilização do formato em documentos nado-digitais ou que tenham sido objeto de OCR:

- PDF/A-1a: Nível A conformidade na parte 1 – tem como objetivo garantir a reprodução fiável da aparência visual do documento e que o conteúdo pode ser pesquisado e reaproveitado; salvaguarda a estrutura lógica e semântica do documento e a sequência do texto (por exemplo o ajustamento do texto em écrans de dimensões diferentes);
- PDF/A-1b: Nível B conformidade na parte 1 – tem como objetivo garantir a reprodução fiável da aparência visual do documento, não considera a coerência textual dos mesmos.

Tabela 1 – A norma ISO 19.005

Norma	Designação		Data	Versão PDF (de suporte)
ISO 19.005-1	PDF/A-1	Use of PDF 1.4 (PDF/A-1)	2005	PDF 1.4 (Adobe Systems, PDF Reference third edition, 2001)
ISO 19.005-2	PDF/A-2	Use of ISO 32.000-1 (PDF/A-2)	2011	PDF 1.7 (ISO 32.000-1:2008)
ISO 19.005-3	PDF/A-3	Use of ISO 32.000-1 with support for embedded files (PDF/A-3)	2012	PDF 1.7 (ISO 32.000-1:2008)

Ao PDF/A-1 seguiu-se o PDF/A-2 (19.005-2), com base na Referência PDF 1.7 e procurando a conformidade com a ISO 32.000. Este formato admite a compressão de imagem JPEG2000 e a inclusão de multimédia, suportando efeitos de transparência e

camadas, a incorporação de fontes *OpenType*, assinaturas digitais em conformidade com o *PDF Advanced Electronic Signatures* (PAdES) e a opção de embeber ficheiros PDF/A com vista ao armazenamento de conjuntos de documentos num único ficheiro.

O PDF/A-2 deve ser compatível, ou seja, todos os documentos PDF/A-1 válidos também devem ser compatíveis com PDF/A-2. No entanto, ficheiros compatíveis com PDF/A-2 podem não ser necessariamente compatíveis com PDF/A-1. Esta parte 2, define três níveis de conformidade: PDF/A-2a, PDF/A-2b e o PDF/A-2u, um novo nível de conformidade. O PDF/A-2u representa o nível B de conformidade (PDF/A-2b) com a exigência adicional de que todo o texto no documento tem mapeamento Unicode.

Com o mais recente PDF/A-3 (19.005-3) visam-se os documentos dinâmicos e o formato passa a suportar a possibilidade de embeber em ficheiros PDF/A formatos diversificados como XML, CSV, CAD, documentos de processamento de texto, folhas de cálculo (como objetos arquivados completos).

Em termos gerais, o formato PDF/A não constitui um sistema ou estratégia de arquivo/preservação, nem exclui formatos considerados como “de arquivo”, nomeadamente o TIFF, identificando-se como principal contributo da norma ISO 19.005 a determinação de um “perfil” perspetivado para o longo prazo. A implementação bem-sucedida do PDF/A não resulta apenas da utilização do formato mas de uma atuação integrada que envolve: políticas, estratégias e procedimentos de gestão da informação; especificação de requisitos e condições adicionais; e processos de garantia de qualidade.

O PDF/X não se apresenta como uma alternativa ao PDF, sendo um subconjunto focalizado de PDF especificamente concebido para o intercâmbio de dados de pré-impressão confiável. É uma norma de aplicação, bem como uma norma de formato de ficheiros, ou seja, esta define como as aplicações que criam e lêem ficheiros PDF/X se devem comportar. Os principais objetivos do PDF/ X são:

- melhorar a cor e a correspondências de conteúdo de prova para prova, prova para a impressão, e impressão para impressão;
- reduzir os erros de processamento em prova e pré-impressão;
- permitir a rápida, eficaz e automatizável pré-visualização de ficheiros no momento da receção do cliente;
- reduzir a complexidade e o custo da educação do cliente (BAILEY, 2005).

Os formatos PDF/X são projetados para serem amplamente aplicáveis em diversos setores e áreas geográficas ligados à indústria da impressão. Assim, constituem uma base muito forte para o desenvolvimento de especificações personalizadas, mais concretamente para um sector específico.

A combinação de duas divisões conduziu à criação de vários formatos PDF/X:

- PDF/X-1a. Aborda os intercâmbios cegos em que todos os ficheiros devem ser entregues em CMYK, sem RGB ou dispositivo de dados (com gestão de cores) independentes. Esta é uma exigência comum em muitas áreas e setores de impressão.
- PDF/X-3. É um superconjunto do PDF/X-1a; um ficheiro PDF/X-1a atende a todos os requisitos técnicos de um PDF/X-3. Os padrões PDF/X-3 e PDF/X-2 são claros sobre a forma como uma prova ou um dispositivo *placa-setter* deve agir sobre as cores num ficheiro. Todas as ferramentas concebidas para ler PDF/X-3 também devem ser capazes de ler ficheiros PDF/X-1a.
- PDF/X-2. Foi concebido para abordar os intercâmbios em que há mais “discussão” entre o fornecedor e o recetor do ficheiro (BAILEY, 2005).

Por sua vez, o PDF/E – norma ISO 24.517-1:2008⁷ – é um subconjunto do PDF, projetado para ser um formato de troca aberta e neutra para documentação técnica e de engenharia para a criação de documentos utilizados na construção de *workflows* (BAILEY, 2005). A norma específica como o PDF/E deve ser utilizado para a criação de documentos em *workflows* de engenharia. Os principais benefícios apontados incluem:

- Reduzir os requisitos de *software* caro e proprietário;
- Menores custos de armazenamento e taxas (*vs.* papel);
- Troca confiável em vários aplicativos e plataformas;
- Ser independente.

A norma não define um método para a criação ou conversão de papel ou documentos eletrônicos para o formato PDF/E tendo sido criada para atender às necessidades das organizações que precisam de uma forma confiável para criar, trocar e rever documentação de engenharia, no entanto, a primeira parte da norma não trata de 3D, vídeo ou outro conteúdo dinâmico.

b. A utilização do formato PDF⁸

De acordo com o documento publicado pela Library of Congress, *Recommended Format Specifications*⁹, para que exista uma boa preservação dos ficheiros digitais existem várias características preferenciais que devem ser tidas em conta como boas práticas:

⁷ ISO 24.517-1:2008 – *Document management : engineering document format using PDF. Part 1: Use of PDF 1.6 (PDF/E-1)*.

⁸ Cf. Fernandes, Brandão e Costa, 2010; Brandão, 2010; CMP, 2012; Sousa, 2013; Oliveira, 2014.

⁹ Cf. a nova versão em: *Recommended Format Specifications*. [Consult. 10 out. 2014]. Disponível em <http://www.loc.gov/preservation/resources/rfs/>.

A. Caraterísticas técnicas

1. Maior resolução disponível, não redimensionada ou interpolados
2. Maior profundidade de *bits* disponível, 16 *bits* por canal, se disponível
3. Especificação do espaço de cor usado na versão publicada
4. Não-comprimidas
5. Não estratificadas
6. Melhor versão disponível, clareza total (por exemplo, maior resolução, maior a profundidade de *bits*, sem compressão).

B. Formatos, por ordem de preferência

1. TIFF (não comprimido)
2. JPEG2000 (sem perdas)
3. PNG
4. JPEG / JFIF
5. DNG (Negativo Digital)
6. JPEG2000 (com perdas)
7. TIFF (comprimido)
8. BMP
9. GIF

C. Metainformação

1. Se for suportado pelo formato, incluir a seguinte metainformação: título, autor, data de criação, local de publicação, editor / produtor / distribuidor, informações de contacto. Incluir, se disponível: língua de trabalho, outros identificadores relevantes (por exemplo, DOI, LCCN, etc.), descrição de assunto, resumos, informações de produção chave ou referência a cada campo de dados e metainformação técnica.

D. Medidas de carácter tecnológico

1. Os ficheiros não devem conter nenhuma medida que controle o acesso ou uso do trabalho digital (tais como gestão de direitos digitais ou criptografia).

Este documento menciona, ainda, que apesar das descrições acima serem as ações preferenciais para a preservação de ficheiros digitais, existem também algumas medidas que são práticas igualmente aceitáveis no tratamento destes ficheiros:

A. Formatos

1. TIFF (em formato Planar)
2. PhotoShop
3. Camera RAW
4. JPEG 2000 Parte 2

5. FlashPix
 6. Computação Gráfica Metafile (CGM, WebCGM)
 7. PostScript encapsulado (EPS)
- B. Compressão
1. Taxas de compressão mais baixas preferenciais
 2. *Discrete Wavelet Transform* (DWT) preferido a *Discrete Cosine Transform* (DCT)
- C. Medidas de caráter tecnológico
1. Os ficheiros não devem conter nenhuma medida que controle o acesso ou utilização da obra digital (como gestão de direitos digitais ou criptografia)

Tomando como referência o trabalho desenvolvido na CMP, a informação produzida e transferida para formato digital apresenta-se sob a forma de:

- Imagens;
- Documentos nado-digitais (estruturados, semiestruturados e não estruturados em vários formatos);
- Outra informação nado-digital (bases de dados, ...).

A preservação da dimensão lógica da unidade informacional foi equacionada identificando-se os problemas relacionados com a obsolescência dos formatos, a interoperabilidade, assim como o problema dos formatos proprietários. Na Tabela 1 apresentam-se os *formatos normalizados*, isto é, a adotar com vista à preservação no longo prazo.

Porém, nem sempre o formato de preservação é o mais indicado para a fase em que o documento ainda está em tramitação, assim como não pode ser ignorada a necessidade de proceder à captura e gestão da metainformação que documente e assegure o acesso e uso das unidades informacionais digitais no longo prazo.

Tabela 2 - Taxonomia de propriedades significativas (documento interno CMP, 2012)

Classes	Formato transacional	Formato definitivo (preferencial)
Texto estruturado	Word, Excel, OpenOffice, etc.	PDF A
Imagens	Png, Gif, Jpeg, vectorial (dwf), etc.	TIFF, JPEG 100% e DWF X
Base de dados relacionais	Access, Oracle, SQL Server	DBML

No entanto, e como se apresenta na Tabela 2, existem diferentes tipos de metainformação cuja captura e manutenção deve seguir as normas internacionais, que, afirmando-se pela aplicação generalizada, resultam do reconhecimento internacional face a boas práticas desenvolvidas nos diferentes domínios, nomeadamente no setor da Informação/Documentação.

Tabela 3 - Matriz de tipos de metainformação por função e normas
(documento interno CMP, 2012)

Tipo de Metainformação	Função	Normas
Técnica	Preservação; pesquisa	NISO Z39.87
Estrutural	Apresentação, reconstituição do objeto	METS, Dublin Core
Descritiva	Acesso	EAD; ISAD(G); ISAAR(CPF)
Preservação	Autenticidade	PREMIS

Neste contexto, e partindo do modelo OAIS (*Open Archival Information System*), foi adotado o conceito de “pacote” que agrega a informação e a correspondente metainformação: *Pacote de Informação para Submissão* (PIS/SIP), *Pacote de Informação Arquivada* (PIA/AIP) e *Pacote de Informação para Disseminação* (PID/DIP).

Na CMP todo o projeto relativo à estruturação da Unidade Central de Digitalização (UCD) teve em consideração estes problemas e possíveis soluções, passando quer pela análise exaustiva da produção informacional que iria ser objeto de digitalização e consequente definição de perfis de digitalização, quer pela adoção do PDF/A, quer pela captura/produção e registo de metainformação e implementação da construção dos pacotes previstos no modelo conceptual OAIS.

No caso da UCD foi dada particular atenção à norma NISO Z39.87 (*Technical Metadata for Digital Still Images*), e à sua versão MIX (em XML), que se apresenta como um esquema que define um conjunto normalizado de elementos de metainformação para imagens digitais, cuja organização é compatível com o modelo de dados PREMIS (*PReservation Metadata Implementation Strategies*), direcionado à metainformação de preservação.

Assim, e no âmbito do processo de transferência de suporte de informação que iria integrar um processo administrativo e respetivo *workflow*, foram definidos, entre outros aspetos, o formato de ficheiro a utilizar, a metainformação a capturar e o processo de autenticação.

Os campos de metainformação recolhidos pelas máquinas digitalizadoras incluem:

- Comprimento de imagem;
- Configurações de digitalização;
- Data de produção;
- Dimensão;
- Dispositivo de captura;
- Esquema de cor;
- Formato;
- Identificador;
- Indivíduo produtor;
- Largura de imagem;
- Modelo de dispositivo de captura;
- N.º de pixéis;
- Nome de fabricante;
- Nome do *software*;
- Número de componentes digitais;
- Plataforma tecnológica;
- Profundidade de *bits*;
- Resolução espacial;
- Sistema Operativo;
- *Software* de captura;
- Versão de *software* de captura;
- Versão do *software*.

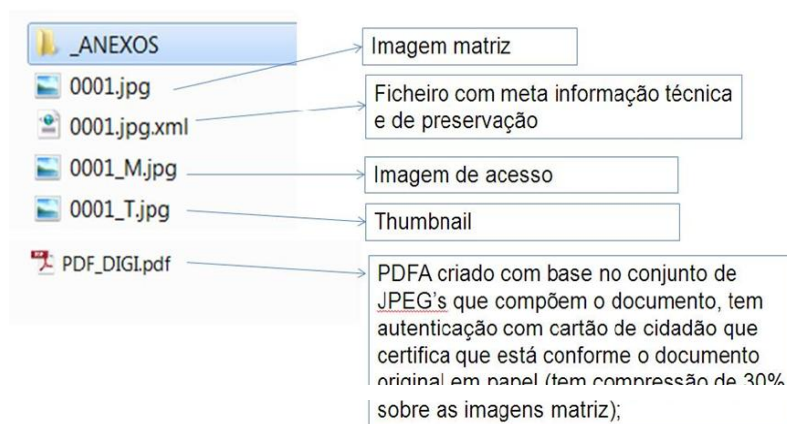
Quanto às características da digitalização ao nível da UCD foram especificadas as seguintes:

- JPEG a 100% (sem compressão);
- 24 *bits* de profundidade;
- 200 Dpi's de resolução;
- a cores;
- captura automática de metainformação técnica para um ficheiro XML, sendo que os campos registados estão de acordo com a norma NISO Z39.87.

Foi, ainda, adotado o formato PDF/A para todos os documentos digitalizados pela UCD. Inicialmente são gerados ficheiros com o formato JPEG 100%, sendo a partir deste gerado o formato PDF/A, com o nível de conformidade 1-b.

Como representado na Figura 5, são guardados vários formatos do mesmo ficheiro, cada um com características e funções diferentes, por forma a responder quer a requisitos de preservação, quer para visualização e uso interno e externo.

Fig. 5 - Formatos de ficheiros armazenados no repositório digital da CMP



O processo envolve a produção de um *formato transacional* e, quando o documento é arquivado, ocorre a normalização com a sua transformação no formato definido para a preservação da informação a longo prazo.

A política e processos subjacentes à transferência de suporte constituirão, naturalmente, parte integrante da *Política de Gestão da Informação* em desenvolvimento na organização, ao abrigo da qual será, por sua vez, especificado o *Plano de Preservação e Segurança da Informação*.

3. A produção informacional via email

O uso do @ na sociedade contemporânea é de imediato associado ao *email* (correio eletrónico), e este, por sua vez, ao ato de compor, enviar e receber mensagens através das Tecnologias de Comunicação e Informação (TIC).

De facto, trata-se de um novo meio de comunicação, muito marcado pela informalidade da mensagem/discurso, que rapidamente se arreigou no quotidiano de pessoas, instituições e organizações, substituindo progressivamente as tradicionais vias de comunicação, como o correio postal, fax, telefone, etc.

A velocidade da utilização do *email* nas rotinas diárias não foi, no entanto, acompanhada pela consciencialização da sua importância como “produto informacional”. Se o ofício recebido em suporte papel continuava a ser objeto de registo, organização, descrição, instalação e armazenamento sob controlo, para posterior recuperação e uso, o *email* recebido mantinha o cunho informal e um tratamento aligeirado e confinado ao âmbito pessoal do recetor, não se diferenciando, de forma substancial, se, em termos

comportamentais, esse recetor se encontrava em contexto pessoal ou no contexto das suas atividades profissionais ou cívicas.

Não obstante, ao nível institucional e organizacional vem ocorrendo nos últimos anos a lenta mas progressiva consciencialização da sua importância como matéria-prima, evidência e suporte para as respetivas atividades, emergindo a necessidade de atribuir ao “*email*”/mensagem recebida por via digital, uma componente formal que assegure a manutenção continuada da relevância informacional que possui, bem como os atributos de autenticidade, integridade, fidedignidade, confidencialidade, inteligibilidade e usabilidade exigidos à informação que integra o sistema de informação, recurso de gestão e memória de indivíduos, instituições e organizações.

Um processo que, todavia, está longe de ser linear e que se confronta com consideráveis dificuldades. Desde logo ao nível da dissociação do ato do foro pessoal, daquele que respeita ao foro institucional/organizacional em que se insere o “agente”, até às dificuldades inerentes ao uso da tecnologia e à rápida obsolescência que lhe é inerente, a par da crescente perceção da complexidade decorrente da pluridimensionalidade que envolve a unidade informacional recebida ou expedida via *email* e que, de facto, integra um conjunto de componentes informacionais e meta-informacionais digitais que, só em conjunto, garantem a unidade de sentido que constitui, efetivamente, a unidade de informação.

3.1. *Em que consiste e como funciona o email?*

Electronic Mail, *e-mail*, *email* ou *correio eletrónico* designam, desde logo, um serviço que permite compor, enviar e receber mensagens através de redes e sistemas eletrónicos de comunicação.

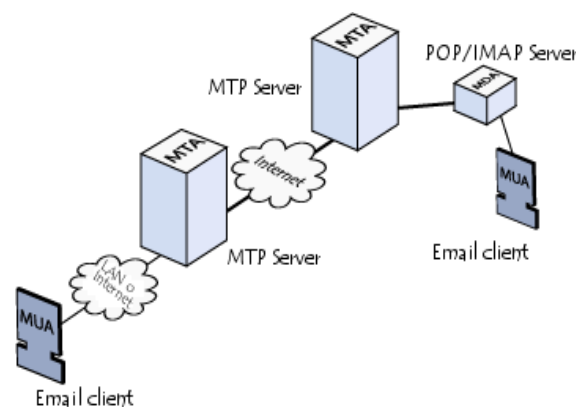
Uma mensagem de *email* é estruturada por duas componentes essenciais, o cabeçalho da mensagem e o corpo da mensagem, constituído pelo conteúdo do *email*, a que se acrescem os anexos. O cabeçalho da mensagem contém informação de controlo, incluindo, como elementos básicos, o endereço de correio eletrónico de um remetente e um ou mais endereços de destinatários.

Quando um *email* é enviado, a mensagem é encaminhada de servidor para servidor, para o servidor de *email* do destinatário. Mais precisamente, a mensagem é enviada ao servidor de correio encarregado de transportar *emails*, chamado MTA (*Mail Transport Agent*), para o MTA do destinatário. Este entrega o *email* para o servidor de correio recebido, chamado de MDA (*Mail Delivery Agent*), que o armazena, aguardando que o utilizador o aceite. Na Internet, os MTA comunicam entre si utilizando o protocolo SMTP,

e, assim, são chamados logicamente servidores SMTP (*Service Mail Transfer Protocol*). São dois os principais protocolos utilizados para a recuperação de *email* do MDA:

- POP3 (*Post Office Protocol*), o mais antigo e usado para a recuperação de *email* e que, em certos casos, deixa uma cópia no servidor.
- IMAP (*Internet Message Access Protocol*), que é usado para coordenar o estado do *email* (ler, apagar, mover) em vários clientes de email. Com o IMAP, é guardada no servidor uma cópia de cada mensagem, de modo a que a tarefa de sincronização possa ser concluída.

Fig. 6 – Processo de envio/recepção de um *email*



Os MTA funcionam como uma estação de correios (a área de triagem e o carteiro, que lidam com o transporte de mensagens), enquanto o MDA funciona como caixa de correio, que armazena mensagens (tanto quanto o seu volume permita) até ao momento em que os destinatários verificam a caixa. Isto significa que não é necessário que os destinatários estejam ligados para lhes serem enviados *emails*. Para manter a confidencialidade dos *emails* de todos os utilizadores, o MDA é protegido por um nome de utilizador (*login*) e por uma senha (*password*).

A recuperação de *email* é efetuada através de um programa de *software* designado por MUA (*Mail User Agent*). Quando o MUA é instalado no sistema operativo do utilizador, é denominado de cliente de *email* (como o *Mozilla Thunderbird* ou o *Microsoft Outlook*), quando é uma *interface web* usada para interagir com o servidor de mensagens recebidas, é conhecido por *webmail*.

3.2. A política de gestão do email

Inserida na *Política de Gestão da Informação* da organização, terá, pois, que ser considerada uma *Política de Gestão de Email*. É certo que esta já vem sendo adotada por várias organizações, apesar de maioritariamente ativada pela via informática e ainda muito confinada a aspetos como a segurança e a confidencialidade, não perspetivando, por exemplo, o processamento, armazenamento e preservação dos *emails* para acesso continuado no longo prazo.

A definição de políticas tende a resultar do impacto da utilização das TIC em termos dos potenciais prejuízos que podem trazer aos colaboradores no contexto organizacional. Foi o caso do Departamento de Administração do Estado do Wisconsin, nos Estados Unidos, no âmbito do qual foi desenvolvida uma *Política de Uso de Internet e Email* com vista a incentivar o uso adequado da Internet e do *email* por parte dos funcionários desse mesmo departamento, minimizando os riscos de utilização de ambas as ferramentas.

As diretrizes incentivam todos os funcionários a usar o bom senso quando recorrem à Internet do Departamento ou acedem ao *email*. Embora essas diretrizes definam como os funcionários podem ou não podem utilizar a Internet e o *email*, estas não conseguem cobrir todas as situações possíveis, surgindo aqui o apelo ao bom senso e sentido profissional de todas as partes. A título de exemplo, enquanto o uso pessoal limitado da Internet e do email é permitido, o acesso excessivo a *sites* não relacionados com trabalho não é permitido (caso do desporto, férias e planeamento de viagens, produtos de consumo e entretenimento).

De acordo com Prom (2011) existem três passos básicos que as instituições que visam implementar projetos de preservação de *email* devem realizar: a definição de políticas, a escolha de ferramentas apropriadas e a sua implementação à luz de fatores ambientais locais e recursos disponíveis. Desta forma, as instituições devem começar por definir as *Políticas de Gestão* e de *Preservação de email*.

Segundo Prom (2011), as políticas de *email* devem esboçar:

1. Um compromisso institucional para a preservação de *email* e ações específicas que serão tomadas, assim como apoiar procedimentos;
2. As expectativas do utilizador final, as responsabilidades e os direitos sobre o acesso, uso, privacidade e controlo das unidades informacionais que integram um *email*.

As *Políticas* devem identificar os “tipos” de *email* que são efetivamente pertinentes e relevantes para a atividade, gestão e memória organizacional/institucional.

Por sua vez, os *Procedimentos* vão definir como os sistemas apoiam a política, e como os utilizadores interagem com os sistemas, permitindo à organização gerir o *email* durante todo o ciclo de vida.

Definidas as políticas adequadas, as instituições devem selecionar e implementar ferramentas apropriadas que suportem as políticas. A implantação tem de ser realizada em colaboração com os utilizadores de *email*, os gestores de informação e os gestores de TI. As instituições devem ter particular atenção em não impor inadvertidamente configurações de preservação que possam dificultar o envolvimento dos utilizadores. Isto pode ser conseguido fornecendo espaço de armazenamento adequado e evitando configurações de auto-eliminação.

Quanto aos utilizadores, estes devem, em primeiro lugar, familiarizar-se com o funcionamento do serviço de *email* que utilizam, bem como os “clientes” que usam para operarem. Entendida a estrutura das contas que possuem e a sua configuração para gerir o *email* de uma forma mais eficaz, os utilizadores podem usar ferramentas de *backup* simples, (como o aplicativo gratuito *MailStore Home* ou a ferramenta de *backup Time Machine* da *Apple*), para garantir a salvaguarda de uma cópia dos seus *emails* num local seguro, separado das cópias utilizadas diretamente pela aplicação informática.

Estes podem, ainda, utilizar um serviço baseado em nuvem (*cloud*) para fazer *backup* dos *emails*. Porém, devem considerar cuidadosamente os termos do serviço antes de o subscreverem. Alguns serviços de *backup online*, (tais como Carbonite e Mozy), realizam um *backup* automático de tudo que exista nas pastas de dados de um computador pessoal, incluindo cópias locais de mensagens de *email*, mantidos em qualquer formato que a “máquina” cliente do utilizador utilize para armazenar.

Segundo a Osterman Research (2010) é importante perceber que fazer *backup* e arquivar *emails* são duas práticas cruciais embora não sejam substitutas uma da outra, ou seja, a diferença fundamental entre *backup* e *arquivo*, é que o primeiro é concebido como uma solução de curto prazo, tática, com foco na informação que é importante para restaurar o bom funcionamento de um servidor, enquanto o arquivo é uma solução mais estratégica e de longo prazo, focada em informação e metainformação que é importante para manter a integridade de toda a informação gerada ou recebida por uma organização ou instituição.

3.3. A produção/captura de email

Os UK National Archives (2011), um serviço especializado de arquivo de âmbito nacional do Reino Unido, produziram um documento que tem como objetivo estabelecer as *Orientações para a Gestão e Preservação de emails* (UK National Archives - *Email*

Management and Preservation Guidelines), ou seja, fornecer um aconselhamento básico, para que o *email* seja gerido de uma forma que facilite a sua preservação a longo prazo e / ou eventual transferência para um *repositório de arquivo*. Segundo o UK National Archives (2011), *Email programs are one common technology that people use to communicate with others, to share information and to record decisions. The semi-formal nature of an email message means that it provides greater reliability as evidence than, say, a text message or a conversation*. Por esta razão, os *emails* devem ser geridos de forma adequada depois de terem sido enviados ou recebidos.

Assim, e devido ao facto de cada programa de *email* guardar as mensagens de forma diferente, o Guia referenciado aplica quatro princípios gerais para que cada pessoa possa gerir o seu correio eletrónico com vista a uma preservação eficaz a longo prazo.

- Usar as funções de "arquivo" do *email* com cautela: usar a função de "auto-arquivo" ou "arquivo" de um programa de *email* com cuidado, sendo necessário algum tempo para compreender e configurar estes recursos;
- Saber onde é armazenada a mensagem de *email* produzida ou recebida, pois os programas e servidores de *email* podem gravar ficheiros em vários locais do disco rígido;
- Organizar as suas mensagens para preservação, existindo muitas formas de organizar *email* "não-corrente" para que seja acessível de uma forma otimizada;
- Independentemente do método escolhido, o sistema deve ser fácil de usar e deve preservar as mensagens que têm valor a longo prazo;
- Usar as pastas locais de forma inteligente: às vezes é necessário armazenar as mensagens num computador local, em vez de num servidor central, devido às quotas impostas ou outro tipo de restrições.

Por seu lado, os National Archives and Records Administration (NARA, USA), uma entidade congénere da referida, publicaram em 2011 o artigo *Guidance concerning the use of e-mail archiving applications to store e-mail* onde são apresentadas as vantagens e desvantagens das aplicações de arquivo de *email*, no contexto de áreas referenciadas como de "Gestão Documental" e de "Gestão de Arquivos". Este artigo fornece aos órgãos federais linhas orientadoras relativas ao uso de aplicações de arquivo de *email* e tecnologias semelhantes para a respetiva gestão. Salienta que, apesar dessas aplicações de arquivo de *email* poderem oferecer benefícios para os processos desenvolvidos por um determinado organismo, estes podem ser limitados na sua capacidade de manter e organizar a informação de acordo com os requisitos legais de gestão, regulamentos e políticas.

Tendo cada aplicação diferentes características e diferentes pontos fortes e dependendo da própria organização e dos seus objetivos comerciais, as aplicações de arquivo de *email* podem oferecer vários benefícios:

- Um armazenamento mais eficiente de *email*, pois é movido, de uma rede distribuída de servidores, aplicativos de *desktop* e outros locais a serem geridos, para um único lugar;
- A capacidade de pesquisa eletrónica avançada de conteúdo que pode ser pertinente para, por exemplo, uma intimação judicial, um pedido de acesso à informação, ou algum propósito semelhante;
- Auxílio no *backup* e na recuperação de desastres.

De acordo com a iniciativa desenvolvida pelo Governo do Canadá deverá ser acompanhado todo o ciclo de vida dos registos de *email* a saber:

- Mensagens de *email* relativas aos negócios/atividades que devem ser mantidas pela instituição:
 - Mensagens de *email* criadas, recolhidas, recebidas ou transmitidas durante o curso normal das atividades da organização são registos do Governo do Canadá e por isso devem ser mantidas para garantir a integridade da memória da governação.
 - Mensagens de *email* cujo conteúdo seja de natureza pessoal, não são registos do Governo do Canadá, os exemplos incluem mensagens de email sobre informações pessoais de um funcionário, emails relativos a passatempos, publicidade não solicitada, etc. Essas mensagens devem ser apagadas uma vez que não são úteis.
- Mensagens de *email* que devem permanecer intactas:
 - Sempre as que mensagens eletrónicas e os seus anexos dizem respeito às atividades da instituição, estas devem permanecer intatas em termos da sua estrutura (*layout* ou formato e *links* para os anexos e documentos relacionados), conteúdo (a informação contida na mensagem) e contexto (informações relativas ao remetente e destinatários, bem como qualquer informação de cabeçalho e os dados de transmissão, tais como data e hora).
- Mensagens de *email* devem ser capturadas num sistema informacional reconhecível:
 - Mensagens eletrónicas relacionadas com o Governo do Canadá devem ser arquivadas no sistema de informação de acordo com as práticas de gestão da informação da instituição.

- Mensagens de *email* devem ser geridas de forma eficiente e eficaz:
 - Mensagens eletrónicas devem ser geridas de acordo com as boas práticas de gestão de informação/registos, a fim de preservar a integridade desta, atender às necessidades de negócios da instituição e cumprir com as responsabilidades.
 - Boas práticas de gestão de informação direcionada ao ciclo de vida do registo criado, da fase de Planeamento à de Avaliação, irão garantir que as mensagens eletrónicas permanecem acessíveis, ou seja, recuperáveis e legíveis, ao longo do tempo.
- A gestão de *email* deve ser apoiada por políticas corporativas, diretrizes e procedimentos:
 - As instituições governamentais devem garantir que as mensagens de *email* e os respetivos anexos permaneçam acessíveis no arquivo da instituição durante a sua vida útil até à sua eliminação final. A elaboração de políticas institucionais, diretrizes e procedimentos devem abarcar aspetos como:
 - Legislação e regulamentos específicos da instituição;
 - Gestão do sistema de *email* e responsabilidade pela informação contida nas mensagens de email enviadas e recebidas através do sistema;
 - Uso do sistema pelos funcionários para assuntos pessoais ou privados;
 - Responsabilidades para a gestão de mensagens eletrónicas, incluindo a identificação do centro responsável pela gestão do arquivo da instituição;
 - Responsabilidades e diretrizes para a retenção e eliminação dos registos de *email* e o papel do utilizador nesse processo.
 - Acesso e segurança de mensagens eletrónicas.

3.4. A Avaliação e a política de retenção

Em rigor, e numa perspetiva sistémica e integrada da gestão da informação, o *email* é apenas um novo meio de comunicação da informação que constitui a mensagem a comunicar. Como tal, essa informação produzida, recebida e acumulada no contexto da atividade da organização integra o seu sistema de informação e deve ser gerida com a especificidade requerida pelas propriedades que possui no âmbito de uma avaliação que

considere sistemicamente o todo informacional (SILVA; RIBEIRO, 2000; RIBEIRO; SILVA, 2004)¹⁰.

Segundo Celorrio (2010), num sistema de *email* coexistem mensagens irrelevantes, mensagens originais assinadas que devem ser mantidas, mensagens com informação confidencial e possíveis ações judiciais e mensagens com anexos. Desta forma torna-se essencial fazer uma seleção, uma vez que não é de todo aconselhável optar por criar uma série documental que inclua todos os *emails*.

Celorrio (2010) recomenda três passos no processo de seleção:

- Formalizar os *emails* como documentos completos, utilizando os mecanismos de autenticação necessários;
- Incluir as regras de avaliação dentro das mensagens formalizadas pela classificação e as ligações necessárias com as regras de auto-execução;
- Eliminar as mensagens não formalizadas ou não classificadas dentro de 15-30 dias da data de envio ou receção.

Há, no entanto, questões básicas elencadas por Celorrio que se podem colocar como ponto de partida mas que não obstam ao desenvolvimento de um processo de avaliação e seleção no âmbito do processo e/ou série informacional e entre as diferentes séries e processos de negócio, nomeadamente:

- A mensagem está relacionada com as atividades e funções da organização?
- A mensagem contém informação com utilidade imediata e unicamente de curto prazo?
- É um duplicado que chega para conhecimento?
- É um esboço/rascunho a eliminar com a criação da versão final?

Em matéria de gestão de *emails* e documentos eletrónicos em geral, a autora defende que a solução é a implementação de um *sistema de gestão de documentos eletrónico* (EDMS) a menos que se prefira transferi-los para suporte analógico. Esta é uma abordagem básica, ainda em utilização por inúmeras organizações, quer de âmbito público como privado, mas que ignora o ponto de partida fundamental que consiste na constatação de que o que nasce em meio digital só poderá ser efetivamente preservado em meio digital. O próprio sistema de “Gestão Documental” poderá ser contraproducente caso não seja assumido no âmbito de uma política integrada de gestão de informação.

¹⁰ SILVA, Armando Malheiro da; RIBEIRO, Fernanda – A Avaliação em Arquivística: reformulação teórico-prática de uma operação metodológica. *Páginas a&b: arquivos e bibliotecas*. Lisboa. ISSN 0873-5670. 5 (2000) 57-113; RIBEIRO, Fernanda; SILVA, Armando Malheiro da – A Avaliação de informação: uma operação metodológica. *Páginas a&b: arquivos e bibliotecas*. Lisboa. ISSN 0873-5670. 14 (2004) 7-37.

Como se constata, dado o crescimento exponencial da utilização da comunicação via *email* e a facilidade da sua produção através das TIC, o processo de avaliação, seleção e determinação do destino final adquire um papel determinante entre os procedimentos de gestão das mensagens de *email*, influenciando, inclusive, os custos relacionados com a sua posterior gestão e preservação.

A eficaz gestão do *email* não só é necessária como se afirma como um objetivo fundamental no âmbito da gestão da informação. Nos *emails* está contida informação relevante que, em muitos casos, pode ser o único registo criado que documenta uma transação, decisões tomadas ou a determinação ou interpretação política.

As regras devem especificar consistentemente quer a estrutura organizacional e informacional em que se inserem, quer a definição das permissões para quem pode aceder, alterar ou excluir mensagens, anexos e outros registos.

Lundgren (2009), acresce que para atingir este objetivo as organizações devem guiar-se através do processo de desenvolvimento, implementação, monitorização e auditoria do que designa como uma completa política de retenção de *email* utilizando os 10 passos que se seguem:

1. Definir uma política de retenção de *email*;
2. Eliminar as variáveis dificultando a centralização;
3. Instruir os funcionários sobre a política de retenção;
4. Incorporar os regulamentos relevantes para a política de retenção;
5. Identificar os papéis com requisitos de retenção exclusiva;
6. Equilibrar as diretrizes de retenção relacionadas com os custos de TI;
7. Fornecer aos funcionários acesso às mensagens arquivadas;
8. Assegurar que as políticas de retenção possam acomodar os documentos de retenção legal;
9. Verificar que todas as mensagens são arquivadas;
10. Usar a tecnologia para aplicar políticas de retenção.

Numa perspetiva sistémica as mensagens de *email* integram o conjunto informacional que corporiza o sistema de informação e, como tal, serão objeto da operação de avaliação da informação que se deverá orientar pelas propriedades da informação, tal como proposto por Silva e Ribeiro (2000).

3.5. A Organização e o armazenamento de emails

De acordo com a HP (2010), as organizações de todas as dimensões enfrentam o desafio do crescimento explosivo de *email*. Em 2010 já era referenciado que 85% das

comunicações empresariais ocorriam através de *email* e 183 bilhões de mensagens de negócios eram enviadas por dia¹¹.

Por seu lado, os “gestores de armazenamento” tentam conter uma espiral de custos de gestão e armazenamento das caixas de correio e os administradores e advogados corporativos necessitam de uma cada vez mais rápida e rentável forma de pesquisar e recuperar a informação contida em *emails*.

Numa abordagem de âmbito geral, Henriksen *et al.* (2013) elaboraram um guia de boas práticas para a construção de uma infraestrutura de armazenamento digital com vista à preservação de informação digital no longo prazo.

Nesse estudo apresentam uma visão geral dos tipos de suporte de armazenamento e dos sistemas disponíveis para a criação do que consideram “uma boa infraestrutura de armazenamento”. Cada tipo de armazenamento é explicado em termos de suas propriedades e quais as vantagens e os riscos do sistema com vista à preservação a longo prazo.

Identificam, assim, várias infraestruturas de armazenamento podendo estas ser de diferentes tamanhos:

- Um pequeno sistema que pode ser autónomo e manipulado manualmente;
- Um sistema maior mas que muitas vezes precisa de sistemas de gestão automatizados e funções de recuperação.

Alguns tipos de suportes de armazenamento surgem e desaparecem rapidamente, tal como os formatos e o próprio *software*. Um exemplo de *hardware* que já se encontra obsoleto são as fitas DAT (*Digital Audio Tapes*) e também as muito comuns disquetes.

De acordo com os referidos autores as propriedades requeridas em termos de suportes de armazenamento são:

- Fácil de usar: Um pequeno número de meios de comunicação é mais fácil de manusear e manter;
- Fácil de copiar: Quão rápido e fácil é copiar entre dois meios do mesmo tipo? Pode ser feito sem o risco de perda de dados?
- Ampla utilização: Quanto mais difundida é a tecnologia, mais tempo tenderá a sobreviver no mercado;
- Baixo preço: Qual é o custo da obtenção de novas versões do mesmo suporte? Quanto espaço de armazenamento pode obter com a verba

¹¹ Dados referentes ao ano de 2010.

disponível e quanto será necessário no futuro? Pensando a longo prazo, isto é, o que no arranque pode ser uma solução barata, no longo prazo pode tornar-se caro.

- Durabilidade: Qual é a expectativa de vida do suporte, antes de ocorrer qualquer risco de erro? Certificar se é um formato estável e se tem um ciclo de vida bem conhecido.

Os autores apresentam, ainda, algumas vantagens e desvantagens por cada tipo de suporte de armazenamento que se sistematizam na Tabela 4.

Tabela 4 – Tipos de suporte de armazenamento (adapt. de HENRIKSEN, *et al.*, 2013)

Tipo de Suporte	Vantagens	Desvantagens
Fitas de Áudio (DAT)	<p>Baixo preço.</p> <p>Escalabilidade.</p> <p>Poupança de energia.</p> <p>Taxas relativamente elevadas de transferência.</p> <p>Baixas taxas de erro em relação a unidades de disco rígido.</p> <p>Suporta compressão e encriptação.</p>	<p>Escreve de forma linear, o que significa que o acesso excessivo não é recomendado, pois a fita vai sofrer um maior desgaste.</p> <p>Os tempos de acesso são longos.</p> <p>Movimentação manual a menos quando conectados em carregadores automáticos.</p> <p>Pode tornar-se uma solução cara, se são necessários carregadores automáticos robóticos.</p> <p>O equipamento leitor / gravador não é compatível com as futuras gerações e precisa ser atualizado.</p>
Discos rígidos	<p>Acesso aleatório e rápido.</p> <p>Contém índice de ficheiros gerido por sistemas de índices de ficheiros.</p> <p>Múltiplos ficheiros podem ser abertos e utilizados ao mesmo tempo por vários utilizadores.</p> <p>Portátil entre plataformas e sistemas operacionais.</p> <p>Escalável.</p>	<p>Para dimensionar o conteúdo pode ser uma solução cara.</p> <p>Desperdiça energia.</p> <p>Sistemas SATA HDD têm provado ter taxas de erro mais elevadas do que as fitas magnéticas.</p> <p>Vida útil de apenas cinco anos.</p>
Discos de estado sólido (SSD)	<p>Baixo consumo de energia.</p> <p>Tamanho compacto.</p>	<p>Preço elevado.</p> <p>Baixo desempenho na escrita de dados (isso pode ser melhorado com um</p>

	Resistente ao choque. Alto desempenho para acesso a dados aleatórios.	determinado <i>software</i> incorporado). Tempo de vida incerta.
Discos Óticos (<i>Blu-Ray</i>)	Possui uma melhor qualidade de imagens que proporciona um melhor aproveitamento de um vídeo, filme. Contém grande espaço de memória para armazenar dados.	É uma tecnologia cara. Os equipamentos e manutenção deste tipo de tecnologia têm custos elevados.
Armazenamento em nuvem (<i>Cloud</i>)	Custo / benefício. Pode ter redução de custos. Uma menor pressão sobre o departamento de TI para manutenção. Fácil acesso, também fora das instalações da organização. Melhor e mais fácil colaboração – geograficamente.	Segurança. Sem plano de preservação Sem controlo e verificação da integridade dos ficheiros. Ainda não é ideal como uma solução de preservação a longo prazo. Menos controlo sobre os próprios conteúdos.

Incidindo especificamente sobre a gestão de *email*, Houston (2008) analisa a identificação, organização e preservação de email com vista à sua preservação a longo termo.

De acordo com o *Electronic Communication Guidance for University Records* (UNIVERSITY OF WISCONSIN, 2012) deve-se começar com a criação de um sistema normalizado de arquivo (etiquetas/metainformação, ordenação e organização) para armazenar as mensagens e para se poder recuperar e usar a informação eficientemente ou tomar decisões com base nesta.

Deverão ser usados estruturas de nomeação de ficheiros consistentes dada a sua importância para aceder e recuperar informação de uma forma adequada, permitindo, assim, responder a pedidos ou agir no âmbito do fluxo de trabalho diário.

Ainda segundo o *Electronic Communication Guidance for University Records*, (UNIVERSITY OF WISCONSIN, 2012) uma sugestão para a criação de uma estrutura de arquivo é utilizar como referência as políticas de retenção e arquivo de correspondência em papel, o que, na nossa perspetiva deverá ser ponderado não em função unicamente desta série em específico mas no âmbito do processo informacional em que se integra

cada *email*, tal como a restante produção informacional, a série correspondência tenderá a ser uma entre várias.

Desta forma, e uma vez que foi desenvolvida uma estrutura de pastas, que, na nossa perspectiva, tende a coincidir com a estrutura de classificação adotada pela organização, devendo esta ser utilizada e aplicada consistentemente para facilitar o acesso, recuperação e consulta de informação.

Estando a gestão de *email* orientada para o utilizador, cabe, desde logo, ao utilizador/produtor (que identificamos como possuidor) gerir o seu email de forma adequada e de acordo com as políticas da organização, a que acrescentamos a necessidade do seu desenvolvimento sob uma abordagem holística, sistémica e integrada do SIO face à Missão e ação da organização.

A título exemplificativo é apresentado por este guia, um quadro geral de *emails* rececionados que apresenta as seguintes tendências:

- Aproximadamente 50% dos *emails* serão designados como *não-registos*.
- Cerca de 25% será de natureza transitória.
- Cerca de 15% será de natureza rotineira.
- Cerca de 10%, ou menos, será gerido de acordo com a análise resultante do processo de avaliação e seleção da informação.

Desta forma, a chave para a gestão de *email* é excluir os *não-registos* (informação transacional) e gerir quaisquer registos de rotina ou transitórios, segundo a Tabela Geral de Temporalidade aplicável à informação de comunicação organizacional.

Independentemente das diferenças patentes nas abordagens desta fase do ciclo de vida da informação, ressalta-se o foco na operação cada vez mais importante da “Avaliação da Informação” que, na nossa perspectiva, deverá ser objeto de um apurado estudo do sistema de informação e ocorrer cada vez mais no início do ciclo de vida e de gestão da informação.

3.6. A Preservação de emails

Tal como qualquer outro tipo de informação, a preservação da produção informacional comunicada via *email* terá que considerar quer a componente física, quer digital. Como refere Pinto (2010), estando em causa informação produzida em meio digital, a questão física também se coloca, dado que aquela mantém a necessidade de ser registada, uma materialização que ocorre através de plataformas tecnológicas com componente de

hardware e *software* (física e lógica), podendo mesmo ser acompanhada pela impressão da mensagem e/ou anexos em papel.

Na perspectiva da unidade informacional esta apresenta-se como pluridimensional, isto é, possuindo várias dimensões e exigindo cada uma delas uma atenção particular: a dimensão física, a dimensão lógica, a dimensão conceptual e a dimensão essencial (as três primeiras referenciadas e designadas como *multi-layered nature of digital objects* por Thibodeau) (2002).

A preservação da informação digital requer, pois, diferentes abordagens, quer técnicas, quer organizacionais pelo que, abordam-se, de seguida, alguns dos principais problemas e estratégias de preservação a considerar no caso de *emails*.

A preservação de *emails* coloca vários problemas, muitos deles similares à restante informação digital. Apesar de ser mais fácil criar, corrigir e distribuir informação digital, os sistemas de armazenamento são mais frágeis do que os tradicionais. Garantir o acesso a longo prazo da informação armazenada digitalmente é um desafio e, cada vez mais, é vista como uma parte importante da gestão de informação em meio digital.

A preservação envolve a retenção da informação, relativa quer ao “objeto” quer ao conteúdo, sendo de acrescer, em termos de complexidade, a dependência que do “meio” tecnológico que medeia a produção, gestão e acesso à mesma. A constante evolução da tecnologia provoca ciclos de obsolescência extremamente rápidos, provocando uma descontinuidade com a consequente possibilidade de inacessibilidade ao nível das várias dimensões identificadas. Daí que se possa afirmar que os recursos informacionais digitais apresentam mais problemas do que os recursos tradicionais.

Centrando-se na preservação de *emails* a longo prazo, Houston (2008) considera ser necessário questionarmo-nos sobre três vertentes essenciais:

- O *Suporte/plataforma*: o meio de armazenamento é durável o suficiente para manter a sua integridade ao longo do tempo?
- A *Mensagem*: o conteúdo do documento é devidamente preservado?
- A *Metainformação*: existe suficiente informação complementar para contextualizar o documento?

Estes componentes deverão existir em simultaneidade. Caso se perca apenas um desses componentes da mensagem eletrónica, a preservação desta não será realizada de forma adequada.

Tomando como exemplo o armazenamento da informação em meio digital, ter-se-á que referenciar este aspeto da preservação e “automaticamente” pensar nas condições ambientais a que os dispositivos de armazenamento de *hardware* serão expostos, sendo que ao armazená-los corretamente pode-se aumentar a expectativa de vida da informação.

De acordo com Henriksen et al. (2013), o meio ambiente para um armazenamento físico ideal deverá ter as seguintes características:

- Humidade relativa mantida entre 35% -40%.
- Temperatura mantida entre 15º e 21º C (dependendo do tipo de *hardware*).
- Monitorização da temperatura e da humidade relativa.
- Alarme de incêndio.
- Sistema de extinção de incêndio sem recurso a água.
- Acesso restrito à área de armazenamento.
- Existência de filtros de ar.
- Blindagem magnética (especialmente para fitas magnéticas).
- Ter instaladas câmaras de vigilância.
- Possuir fonte energia para *backup* em caso de desastre.
- Na iluminação prever proteção de raios ultravioleta assim como em todas as janelas.
- Excluir condutas de água perto, ou por cima, da área de armazenamento.
- A área de armazenamento não deve ser construída em cimento (aumenta a humidade para quase 100% em casos de incêndio).

Assumindo a Preservação como variável da Gestão da Informação foi referido que aquela ocorre desde que se está a preparar a especificação da plataforma tecnológica que suportará a produção informacional.

A informação produzida em meio digital deverá ser mantida e preservada nesse meio. No entanto, não se pode esquecer que as suas morfologias podem ser várias: textos, bases de dados, imagens (fixas ou em movimento), gravações sonoras, material gráfico, programas informáticos e, entre outros, as mensagens de *email*. Os anexos dos *emails* não têm apenas um tipo de conteúdo, comportando texto, imagens, vídeos, animações. Daí que a grande diversidade de formatos não permita uma solução única de preservação, tornando-se necessária uma estratégia mais ampla para atender, pelo menos, aos tipos de formatos mais utilizados.

A informação em meio digital possui características específicas. Não se tem a percepção direta da informação existente e onde está armazenada. A sua estrutura e conteúdo configuram-se no momento da visualização, é uma estrutura lógica e não física. Há, por isso, uma grande dificuldade em localizar os documentos em meio digital e identificar os procedimentos que lhe estão na origem.

No caso da informação de um *email*, esta tem um armazenamento distribuído, está armazenada em diferentes servidores, possivelmente em diferentes partes do mundo, e pode ser acedida de diversos pontos físicos. Para a sua gestão é necessária a existência de metainformação, que integra a dimensão essencial construída ao longo de todas as outras, possibilitando uma identificação completa e inequívoca, de modo a garantir segurança em todo o seu ciclo de vida (DELGADO; BARBOSA, 2009).

No que respeita ao enquadramento legal, o *email* é afetado por uma série de legislação que carece de definição (direitos de autor, privacidade, marcas registadas, segredos comerciais, questões de importação/exportação, etc.). Essa legislação tem de ser tida em conta aquando da sua preservação. A par das mudanças tecnológicas, há também constantes mudanças no campo dos regulamentos e legislação, o que pode requer mudanças no sistema de gestão da preservação definido.

Temos, ainda, que estar conscientes de que, apesar de poderem ser usados como prova de transações ou comunicações, é possível criar *emails* fraudulentos e depois apresentá-los como algo verídico. Ao arquivar tem que ser garantida a integridade e a autenticidade dos emails que vão ser preservados.

Desta forma e segundo Prom (2011) supondo que o *email* pode ser capturado e guardado, há de facto uma área legal adicional que precisa de atenção aquando do desenvolvimento de um programa de preservação de *email* – nomeadamente leis de direito de autor e de propriedade intelectual. O direito de autor inerente a qualquer dos componentes de uma mensagem de *email* irá afetar o que pode ser feito com a mesma no longo prazo.

No que concerne às estratégias de preservação a desenvolver em meio digital as possibilidades são diversificadas não passando as opções pela adoção de uma única.

Para a preservação de *email* Houston (2008) apresenta três soluções a longo prazo:

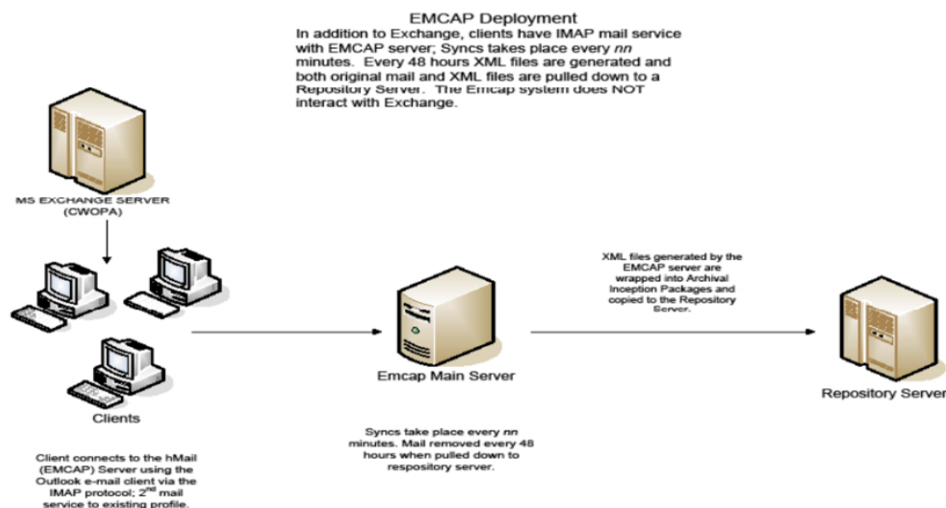
- Imprimir os *emails*:
 - Vantagem: evita os problemas de obsolescência;

- Desvantagem: não é pesquisável nem reutilizável em meio digital, constituindo uma reprodução incompleta do mesmo, o que nasce digital deve permanecer digital;
- Retenção dos *emails* no cliente:
 - Vantagem: armazena documentos criados por aplicativos; uma opção mais fácil para a maioria dos utilizadores, quando bem organizados.
 - Desvantagem: afeta o desempenho do sistema; problemas ao nível do *backup*;
- Armazenar os *emails* num formato neutro:
 - Vantagem: arquivos convertidos para um formato *open source* (TIFF, XML, PDF/A) e armazenados remotamente; reduz/elimina a necessidade de migração ou emulação; oferece opções de *backup*;
 - Desvantagem: a conversão para estes formatos pode envolver um trabalho intensivo.

Nos Estados Unidos, e face aos problemas existentes, os estados do Kentucky e da Carolina do Norte (Arquivos da Carolina do Norte, Kentucky e Pensilvânia) desenvolveram em parceria um projeto de *gestão e preservação de emails*.

Este projeto caracteriza-se por utilizar servidores *email open source*, ter a possibilidade de criar pastas no Servidor de Arquivo (SA), sendo que as mensagens que forem assinaladas como tendo “valor arquivístico” podem facilmente ser copiadas e arquivadas no SA, usando um posto cliente, caracterizando-se, ainda, por ser capaz de copiar a estrutura completa entre uma pasta comum e uma pasta do SA.

Figura 7 – Arquitetura da ferramenta EMCAP (MCANINCH; EUBANK, 2008)



Dele resultou uma ferramenta de gestão e preservação de *emails* (EMCAP) usada para converter o *email* no seu formato nativo para o formato XML por forma a permitir que os utilizadores "arquivem" os seus *emails* numa base sistemática, isto é, de acordo com um plano de classificação, e que fosse capaz de suportar vários tipos de ficheiros, minimizando o suporte das TI.

Assim nasce a ferramenta EMCAP, uma ferramenta *open source*, que permite que o cliente tenha uma estrutura de ficheiros mapeada num servidor e reúne os seus *emails* numa mesma "coleção" (classe ou subclasse numa estrutura). Depois de os dados serem sincronizados, são gerados ficheiros XML das mensagens, que em conjunto com a versão original da mensagem são armazenadas num repositório.

A ferramenta que gera o XML das mensagens contém as seguintes funcionalidades:

- Esquema XML que descreve todos os *emails* numa conta;
- Desenvolvimento de um "esquema comum", com concessão CERP (*Collaborative Electronic Records Project*);
- Análise das informações de cabeçalho no *email*;
- Armazenamento de todos os fluxos de *bits* originais no formato nativo;
- Caso o ficheiro seja externo grava uma síntese da mensagem que é criada com um identificador único;
- Os testes preliminares mostram conversão com 95% de eficácia.

Quanto à *gestão dos anexos* enviados nas mensagens de correio eletrónico, esta ferramenta permite que:

- Os *links* para anexos sejam mantidos em formato nativo e convertidos para Unicode;
- A migração ou conversão podem ser necessárias no futuro para arquivos binários (pdf, doc, etc.);
- O sistema deixa marca no código XML para facilmente se identificar o anexo.

O desenvolvimento de um Esquema XML para Conta de *email* (*E-Mail Account XML schema*) é uma outra possibilidade e viabilizou a preservação de numerosos *emails* relacionados (todo o conteúdo da conta de *email*) num único ficheiro XML.

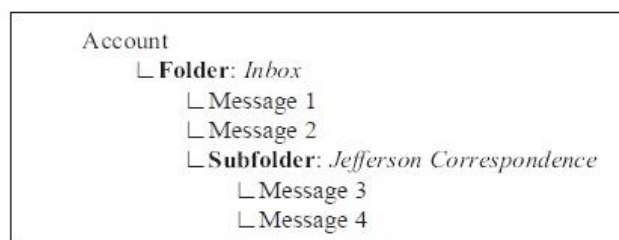
Procurando reter a metainformação inerente a uma conta de *email* e na apresentação das mensagens de *email*, o CERP (*Collaborative Electronic Records Project*) e o EMCAP trabalharam em conjunto para definir um esquema XML que efetivamente capture e preserve as mensagens de *email* de uma forma que estas retenham de forma completa a sua autenticidade e integridade, permitindo aos investigadores, usar uma pesquisa

robusta e estratégias de pesquisa de dados para identificar conteúdo valioso em mensagens individuais, dentro de pastas ou contas (FERRANTE; FUHRIG, 2009).

O esquema utiliza uma estrutura de marcação XML, para incorporar a organização e a estrutura inerente a uma conta de *email*. Além da estrutura organizacional mais básica de uma conta de *email* com uma pasta que contém pelo menos uma mensagem, o esquema precisava de ser robusto o suficiente para lidar com mensagens multiformato, mensagens com anexos e mensagens com mensagens anexadas, e, ao mesmo tempo, capturar a estrutura em múltiplas camadas inerente à organização atribuída pelo proprietário da conta ao *email* nela contido.

De acordo com Ferrante e Fuhrig (2009) a estrutura do esquema da conta de *email* apresenta as mensagens de correio eletrónico nas pastas que os contêm, como é parcialmente ilustrado na seguinte figura:

Fig. 8 – Estrutura parcial de uma conta de *email* preservada (FERRANTE; FUHRIG, 2009)



Este esquema suporta os elementos definidos na RFC 2822 (*Internet Message Format* - norma da Internet para mensagens) para mensagens em conta de email. Assim, os componentes preservados de uma mensagem de *email* prolongam-se para além do conjunto limitado de elementos visíveis por um utilizador típico.

O esquema suporta a incorporação de anexos de mensagens de *email* no ficheiro da conta preservada. Quando isso ocorre o anexo incorporado é mantido dentro da mensagem. Alternativamente, o esquema permite que um anexo de uma mensagem seja arquivado num ficheiro XML externo à conta de *email*.

O resultado final é a completa concretização da preservação de uma mensagem de *email* na sua totalidade - cabeçalho, mensagem e anexos. Quer seja um *email* de texto simples, sem anexos, ou um email multicorpo com documentos, imagens, vídeos e outros *emails* anexados, este leque de possibilidades é considerada na definição do próprio esquema (FERRANTE; FUHRIG, 2009).

Segundo estes autores, o esquema conta de *email* é distinto:

- no seu paradigma baseado na *conta*;
- na granularidade dos dados capturados;
- no alinhamento com a norma para mensagens de *email* RFC 2822;
- em ter um único ficheiro XML por conta;
- na sua incorporação em duas aplicações (software) de preservação de email desenvolvidas separadamente.

O valor-chave desta abordagem é que as inter-relações das próprias mensagens de *email* são preservadas sem a necessidade de documentação adicional, como a informação já existente na conta.

O esquema em si serve como um meio de validar que uma migração de preservação foi concluída com êxito, quando as contas contêm dezenas de milhares de *emails*, torna-se essencial um meio eficiente de verificar a qualidade dos processos de preservação concluídos (FERRANTE; FUHRIG, 2009).

A adesão ao RFC 2822 fornece uma gama mais abrangente e completa de dados, organizados num formato baseado num padrão que o torna mais acessível. A granularidade da estrutura do esquema facilita a acessibilidade e compreensibilidade das contas de *email* preservadas e das suas mensagens, permitindo que estratégias de busca avançada sejam aplicadas a uma ou mais contas simultaneamente (FERRANTE; FUHRIG, 2009).

Devido à estrutura do esquema, é possível pesquisar em toda a conta e recuperar apenas as mensagens que cumpram os critérios estabelecidos e disponibilizar para posterior visualização pelo utilizador.

No âmbito governamental é de salientar logo no início do séc. XXI um projeto do governo holandês especialmente direcionado para as bases de dados relacionais por estas serem amplamente utilizadas no suporte à atuação do governo holandês.

No âmbito deste projeto destaca-se o *Testbed XMaiL - Digital Preservation Testbed*, desenvolvido entre 2001-2003 e direcionado à investigação da preservação digital a longo prazo de diferentes tipos de documentos, nomeadamente: *emails*, documentos de texto, bases de dados, entre outros. Para o caso das mensagens de correio eletrónico, foi criado um protótipo de uma aplicação informática na qual é personalizado o *Microsoft Outlook* com vista a permitir a comunicação com um servidor central onde é recolhida a metainformação e, posteriormente, as mensagens e a metainformação são convertida e armazenadas em XML.

Mais recentemente, e não apenas centrado na busca de soluções tecnológicas, é de referir a iniciativa do Governo do Canadá, a par de outras que se vão multiplicando por todo o mundo. Esta iniciativa envolve a Biblioteca e Arquivos do Canadá destacando-se a publicação do *Guia* para a Gestão do *email* no governo do Canadá (LIBRARY AND ARCHIVES OF CANADA, 2006).

Este guia aborda questões pertinentes para a gestão do *email*¹², tais como o ciclo de vida dos *records*¹³ (registos) de *email*, algumas definições importantes, os procedimentos e papéis no que toca à gestão e proteção do *email*, bem como uma parte destinada para as questões mais frequentes que surgem. Nele é salientado que a quase totalidade dos *emails* produzidos, recebidos e acumulados pelo governo são “*records*”, isto é, destinam-se à retenção no longo prazo:

“A record is under the control of a government institution when that institution is authorized to grant or deny access to the record, to govern its use and, subject to the approval of the Librarian and Archivist of Canada, to dispose of it. Regarding the question of physical possession, a record held by an institution, whether at headquarters, regional, satellite or other office, either within or outside Canada, is presumed to be under its control unless there is evidence to the contrary. A record held elsewhere on behalf of an institution is also under its control, for example at an employee's home or on business travel.

Since most email messages are records, they must be managed in accordance with all applicable legislation and federal government policies such as the Access to Information and Privacy Acts, the Library and Archives of Canada Act, Treasury Boards Management of Government Information (MGI) Policy and the Government Security Policy” (LIBRARY AND ARCHIVES OF CANADA, 2006).

O objetivo deste guia, é que o *email* seja visto como informação que tem de ser gerida e preservada. Ou seja, os *emails* que são criados, recolhidos, recebidos, transmitidos ou

¹² **Electronic mail** (email) messages are communications, sent or received internally or externally on an electronic mail system, and include any attachments transmitted with the message as well as the associated transmission and receipt data.

¹³ **Record** includes any correspondence, memorandum, book, plan, map, drawing, diagram, pictorial or graphic work, photograph, film, microform, sound recording, videotape, machine readable record, and any other documentary material, and any copy thereof. (National Archives of Canada Act, 1987, Access to Information Act R.S. 1985). Definem, também: **Transitory Records** are those records that are required only for a limited time to ensure the completion of a routine action or the preparation of a subsequent record. Transitory records do not include records required by government institutions or Ministers to control, support, or document the delivery of programs, to carry out operations, to make decisions, or to account for activities of government (Authority for the Destruction of Transitory Records, Library and Archives of Canada 1990) [Consult. 04 mar. 2014].

Disponível em: <http://www.collectionscanada.gc.ca/government/disposition/007007-1016-e.html>.

enviados no curso normal das atividades do governo, refletem as funções, as atividades e decisões, constituindo informação oficial do governo. Desta forma, devem ser geridos ao longo do seu ciclo de vida, sendo que para isso terão de existir iniciativas das instituições governamentais para que se estabeleçam serviços, programas, sistemas de gestão e ações de divulgação de informação de forma sustentável e coerente para ir ao encontro das necessidades dos utilizadores. A *gestão do email* deverá constituir parte integrante das políticas, orientações e processos organizacionais.

No âmbito da indústria de *hardware* e *software* esta é uma área para a oferta de novos produtos, sendo a HP, um exemplo entre outros, quando lança um *software* de arquivo de *email* para o software *Microsoft Exchange*. O objetivo deste é fornecer a possibilidade de retenção a longo prazo, a pesquisa de alta velocidade e a recuperação de mensagens e anexos para, assim, reduzir o impacto dos custos na organização.

Constata-se, assim, a lenta mas progressiva consciencialização que vem ocorrendo nos últimos anos, ao nível institucional e organizacional, da importância do *email* como uma via privilegiada para a comunicação interpessoal e interorganizacional/institucional envolvendo informação que constitui evidência e suporte para as respetivas atividades, emergindo a necessidade de atribuir ao “email”/mensagem recebida por via digital a devida representatividade aos diferentes níveis da gestão assegurando-lhe a manutenção continuada da relevância informacional como informação que integra o sistema de informação da pessoa, instituição ou organização, assumida como recurso de gestão e memória de indivíduos, instituições e organizações.

CONCLUSÃO

Este é um processo evolutivo que, todavia, está longe de ser linear e que se confronta com consideráveis dificuldades, face às quais projetos como o que subjaz ao presente artigo poderão gerar contributos que, mesmo sendo sectoriais e parcelares, ajudarão à análise e reflexão em torno do problema da gestão e preservação da produção informacional no longo prazo.

Como foi possível constatar, ao profissional da informação é exigida uma postura, competências e aptidões que lhe permitam estar à altura de uma gestão da informação cada vez mais exigente e que se desenvolve ao longo de todo o ciclo de vida da informação, seja qual for o suporte material e/ou plataforma tecnológica de registo ou modelo de gestão.

Conclui-se chamando a atenção para o facto de esta constituir apenas uma achega para a reflexão sobre um tópico, dos muitos que o desafio da gestão e preservação da informação digital coloca no topo das prioridades estratégicas de instituições e organizações, mas que contribuirá decisivamente para o desenvolvimento de instrumentos essenciais como o *Plano de Preservação da Informação*, que, para serem efetivamente potenciados, deverão existir no quadro global da gestão estratégica da informação corporizada em *Políticas de Gestão da Informação* qualquer que seja a instituição, organização ou setor de atividade em que estas se insiram, e que, neste caso, teve como contexto o projeto em curso num Município.

Referências bibliográficas

BRANDÃO, Marta Diana Nunes

2010 *Arquitectura de Sistemas de Informação alinhada com a política de gestão de informação das unidades orgânicas na Câmara Municipal do Porto*. Porto: Faculdade de Engenharia, Universidade do Porto, 2010.
Dissertação de Mestrado em Engenharia de Serviços e Gestão. Orientador: António Brito.

CANADÁ. Library and Archives

2006 *Email management in the Government of Canada*. [Consult. 5 fev. 2014]. Disponível em: <http://www.collectionscanada.gc.ca/government/products-services/007002-3008-e.html>

CANDY, Jyoti; KARWAL, Vishav.

How e-mail works? PCTE. [Consult. 30 jan. 2014]. Disponível em: <http://www.slideshare.net/adkpcte/how-email-works>

CELORRIO, Fiona

2010 *El Correo electrónico como documento de archivo*. [Consult. 30 jan. 2014]. Disponível em: http://www.errenteria.net/es/ficheros/40_8909es.pdf

DATA DICTIONARY

2005. *Data dictionary for preservation metadata: final report of the PREMIS Working Group*. [Em linha]. [Consult. 15 Ago 2007]. Disponível em: [www.url:http://www.oclc.org/research/projects/pmwg/premis-final.pdf](http://www.oclc.org/research/projects/pmwg/premis-final.pdf)

DELGADO, Cláudia; BARBOSA, Sílvia

2009 *Preservação da Web e email*.
Trabalho para a unidade curricular “Preservação e Conservação” da Licenciatura em Ciência da Informação, das Faculdades de Letras e de Engenharia da Universidade do Porto. 2009.

FANNING, Betsy

2010 *PDF/ARCHIVE: preserving electronic assets*. Washington, DC : NACRC. AIIM [Consult. 5 fev. 2014]. Disponível em: http://www.nacrc.org/assets/Presentations/LC/2010/pdfa_preserving_electronic_assets.pdf

FERNANDES, Daniela; BRANDÃO, Marta; COSTA, Marta

2010 Desmaterializar para potenciar a informação em rede: o caso da UCD da CMP. In CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 10º, Guimarães, 2010 – *Actas do congresso*. Lisboa: APBAD, 2010. [Consult. 5 fev. 2014]. Disponível em <http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/171/166>.

FERRANTE, Riccardo; FUHRIG, Lynda Schmitz

2009 *Digital preservation : using the email account XML schema*. Washington: Smithsonian Institution Archives, 2009.

HENRIKSEN, Sofie Laier; SEUSKENS, Wiel; WIJERS, Gaby

2013 Best practices for a digital storage infrastructure for the long-term preservation of digital files. *Digitising Contemporary Art*. [Consult. 04 mar. 2014]. Disponível em: http://www.dca-project.eu/images/uploads/banners/DCA_D62_Best_practices_for_a_digital_storage_infrastructure_20130506_Version1.pdf

HOUSTON, Brad

2008 *E-mail and records management: identifying, organizing, and preserving e-mail records*. University of Wisconsin-Milwaukee. [Consult. 28 jan. 2014]. Disponível em: <http://www.slideshare.net/herodotusjr/email-management-1862379>

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION

2005 *ISO 19.005-1:2005: document management: electronic document file format for long term preservation. Part 1 – Use of PDF 1.4 (PDF/A-1)*. [S.l.]: ISO, 2005.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION

2008 *ISO 32.000-1:2008: document management: portable document format. Part 1 – PDF 1.7*. [S.l.] : ISO, 2008.

LUNDGREN, Eric

2009 10 steps to establishing an effective email retention policy: white paper: 10 steps to effective email retention. [Consult. 27 jan. 2014]. Disponível em: http://www.ca.com/us/~media/files/whitepapers/10-steps-email-retention-wp-us_198118.aspx

MCANINCH, Glen; EUBANK, Kelly

2008 *Using EMCAP (Electronic Mail Capture and Preservation) to Tame the E-Tiger*. [Consult. 17 jan. 2014]. Disponível em: http://www.history.ncdcr.gov/SHRAB/ar/emailpreservation/docs/emcap_bpe_2008.pdf

OLIVEIRA, Hugo Azevedo de

2014 *A Preservação da informação: um contributo para a implementação de um arquivo digital certificável no Município do Porto*. Porto: Faculdade de Engenharia da Universidade do Porto, 2014.

Dissertação de Mestrado em Ciência da Informação. Orientadora: Maria Fernanda Martins; coorientadora: Maria Manuela Pinto.

PARADIGM

[20--] [Consult. 4 fev. 2014]. Disponível em: <http://www.paradigm.ac.uk/>

PENNOCK, Maureen

2006 Instalment on curating e-mails: a life-cycle approach to the management and preservation of e-mail messages: digital curation manual. [Consult. 4 fev. 2014]. Disponível em: <http://www.dcc.ac.uk/resource/curation-manual/chapters/curating-e-mails/curating-emails.Pdf>

PINTO, Maria Manuela Gomes de Azevedo

2005 Do "efêmero" ao "sistema de informação": a preservação na era digital. *Páginas a&b: arquivos & bibliotecas*. Lisboa. ISSN 0873-5670. 15 (2005) 63-178.

PINTO, Maria Manuela Gomes de Azevedo

2009 PRESERVMAP: um roteiro da preservação na era digital. Porto: Edições Afrontamento; CETAC.Media, 2009. ISBN 978-972-36-1070-3.

PINTO, Maria Manuela Gomes de Azevedo

2013 Gestão de Documentos e meio digital: um posicionamento urgente e estratégico. In CIANCONI, Regina de Barros; CORDEIRO Rosa Inês de Novais; MARCONDES, Carlos Henrique, org. – Gestão do conhecimento, da informação e de documentos em contextos informacionais. Niterói: PPGCI/UFF, 2013. ISBN 978-85-228-1026-0.

PINTO, Maria Manuela Gomes de Azevedo

2014 Da Preservação de documentos à preservação da informação. In DUARTE, Zeny, org – A Conservação e a restauração de documentos na era pós-custodial. Salvador: EDUFBA. 2014, p. 127-196. ISBN 978-85-232-1240-7.

PINTO, Maria Manuela Gomes de Azevedo; SILVA, Armando Malheiro da

2005 Um Modelo sistémico e integral de gestão da informação nas organizações. In CONTECSI - CONGRESSO INTERNACIONAL DE GESTÃO DA TECNOLOGIA E SISTEMAS DE INFORMAÇÃO, 2º, São Paulo, 2005 – *Actas do congresso*. São Paulo: TECSI-FEA-USP, 2005. [CD-ROM]. Disponível em: <http://ler.letras.up.pt/uploads/ficheiros/3085.pdf>

PORTO. Câmara Municipal

2014 *Documento orientador para a criação do arquivo digital certificável da CMP*. Porto: Câmara Municipal, 2014.

PROM, Christopher J.

2010 *Email management and preservation guidelines*. [Consult. 31 jan. 2014]. Disponível em: <http://e-records.chrisprom.com/recommendations/develop-submissioningest-policies/email-management-and-preservation-advice/>

PROM, Christopher J.

2011 *Preserving email: DPC Technology Watch Report 11*. 01 dezembro 2011.

RIBEIRO, Fernanda; SILVA, Armando Malheiro da

2004 *A Avaliação de informação: uma operação metodológica. Páginas a&b: arquivos e bibliotecas*. Lisboa. ISSN 0873-5670. 14 (2004) 7-37.

SILVA, Armando Malheiro da

2006 *A Informação: da compreensão do fenómeno e construção do objecto científico*. Porto: Edições Afrontamento; CETAC.COM, 2006. ISBN 972-36-0859-6.

SILVA, Armando Malheiro da; RIBEIRO, Fernanda

2000 *A Avaliação em Arquivística: reformulação teórico-prática de uma operação metodológica. Páginas a&b: arquivos e bibliotecas*. Lisboa. ISSN 0873-5670. 5 (2000) 57-113.

SILVA, Armando Malheiro da; RIBEIRO, Fernanda

2002 *Das “ciências” documentais à ciência da informação: ensaio epistemológico para um novo modelo curricular*. Porto: Edições Afrontamento, 2002. ISBN 972-36-0622-4.

SOUSA, Paula Maciel Carvalho de

2013 *Segurança e preservação da informação: um modelo para os municípios*. Porto: Faculdade de Engenharia da Universidade do Porto, 2013. Dissertação de Mestrado em Engenharia de Serviços e Gestão. Orientador: António Brito; coorientadora: Maria Manuela Pinto; orientador na CMP: Alexandre Sousa.

THIBODEAU, Kenneth

2002 *Overview of technological approaches to digital preservation and challenges in coming years*. [Consult. 15 mar. 2014]. Disponível em: <http://www.clir.org/PUBS/reports/pub107/thibodeau.html>>.

UNITED KINGDOM. National Archives

2011 *Email management and preservation guidelines*. [Consult. 18 fev. 2014]. Disponível em: <http://e-records.chrisprom.com/recommendations/develop-submissioningest-policies/email-management-and-preservation-advice/>

UNITED STATES OF AMERICA. National Archives and Records Administration

2011 *Guidance Concerning the use of E-mail Archiving Applications to Store E-mail*. [Consult. 20 fev. 2014]. Disponível em: <http://www.archives.gov/records-mgmt/bulletins/2011/2011-03.html>

UNIVERSITY OF WISCONSIN. Madison Archives & Records Management

2012 *Electronic communication guidance for university records*. [Consult. 10 mar. 2014]. Disponível em: <http://archives.library.wisc.edu/records/bulletins/2012%20Electronic%20Communications%20-%20Final.pdf>

Hugo Oliveira | hugooliveirao7@gmail.com

Mestre em Ciência da Informação – Faculdade de Engenharia e Faculdade de Letras da Universidade do Porto

Maria Manuela Gomes de Azevedo Pinto | mmpinto@letras.up.pt

Faculdade de Letras da Universidade do Porto / CETAC.MEDIA